

A mathematical programming based characterization of Nash equilibria of some constrained stochastic games^{*}

Vikas Vikram Singh · N. Hemachandra

Received: date / Accepted: date

Abstract We consider two classes of constrained finite state-action stochastic games. First, we consider a two player nonzero sum single controller constrained stochastic game with both average and discounted cost criterion. We consider the same type of constraints as in [1], i.e., player 1 has subscription based constraints and player 2, who controls the transition probabilities, has realization based constraints which can also depend on the strategies of player 1. Next, we consider a N -player nonzero sum constrained stochastic game with independent state processes where each player has average cost criterion as discussed in [2]. We show that the stationary Nash equilibria of both classes of constrained games, which exists under strong Slater and irreducibility conditions [3], [2], has one to one correspondence with global minima of certain mathematical programs. In the single controller game if the constraints of player 2 do not depend on the strategies of the player 1, then the mathematical program reduces to the non-convex quadratic program. In two player independent state processes stochastic game if the constraints of a player do not depend on the strategies of another player, then the mathematical program reduces to a non-convex quadratic program. Computational algorithms for finding global minima of non-convex quadratic program exist [4], [5] and hence, one can compute Nash equilibria of these constrained stochastic games. Our results generalize some existing results for zero sum games [1], [6], [7].

^{*} A portion of Section 3 (two player case) has been presented in the 8th International ISDG workshop at University of Padova, Italy on 21-23 July, 2011.

Vikas Vikram Singh
Industrial Engineering and Operations Research, Indian Institute of Technology Bombay,
Mumbai 400076, India
E-mail: vikas_singh@iitb.ac.in

N. Hemachandra
Industrial Engineering and Operations Research, Indian Institute of Technology Bombay,
Mumbai 400076, India
E-mail: nh@iitb.ac.in

Keywords Constrained stochastic game, Occupation measures, Single controller game, Decentralized stochastic game, Nash equilibrium, Mathematical program.

Mathematics Subject Classification (2000) 91A10, 91A15, 90C05, 90C20, 90C26.

1 Introduction

It is well known that there is a substantial relationship between game theory and mathematical programming. While it is well known that equilibrium strategies in two player zero sum matrix games are related to optimal points of certain linear programs, in 1964, Mangasarian and Stone [8] have shown that the Nash equilibria of any two player bimatrix game can be obtained from the global maxima of one quadratic program and this approach can be generalized in case of any finite number of players. Later Filar et al. [9], generalized this idea to the infinite horizon stochastic game with finite state space and finite action spaces of all the players. It has been shown that the stationary Nash equilibria of any N -player stochastic game with discounted criterion are in one to one correspondence with the global minima of a certain mathematical program [9], [10]; so, Nash equilibria of such a stochastic game can be computed via the global minima of one mathematical program. The stochastic games described in [9], [10] can be viewed as centralized stochastic games. In such centralized stochastic games all the players jointly control a single Markov chain and all the players have complete information of the Markov chain's state and for taking decision at any time t each player has information of all the actions previously taken by the players. The review article [11] summarizes various algorithmic aspects of zero sum stochastic games along with algorithms for nonzero sum stochastic games with special structure (single controller, etc.). In particular, two player zero sum single controller stochastic game can be solved by a linear program [12], [10], [13] and the Nash equilibria of the nonzero sum single controller stochastic game can be obtained from the global minima of a quadratic program [14].

Since the seminal work of Lloyd S. Shapley [15], stochastic games have come to constitute an important class of models that can capture game theoretic issues among the decision makers involved, apart from accounting for random evolution of the system. The edited volume by Neyman and Sorin [16] has a nice collection of many articles on stochastic games and their applications. The book by Filar and Vrieze [10] presents stochastic games as a natural multi-player generalization of (single player) Markov decision processes and their applications. Constrained stochastic games are realistic because they can capture bounds on consumption of resources, but, are also difficult to analyze. In [3] N -player centralized constrained stochastic games with both discounted and average cost criterion with finite state and finite action spaces are considered and it is shown that there exists a stationary Nash equilibrium under strong Slater condition (irreducibility assumption is also needed in average case). The

existence of Nash equilibrium for constrained stochastic games when the state space is countable and action spaces are compact metric space is discussed in [17]. The characterization of Nash equilibria for general constrained stochastic games via some mathematical program is not known. To the best of our knowledge there are only some special classes of constrained stochastic games which can be solved as linear programs. We give a brief description of all these classes here. The two player zero sum single controller constrained stochastic game with total expected reward criterion and expected average reward criterion is considered in [18], [6] respectively. In both [18], [6] only the player who controls the transition probabilities has constraints on his expected rewards and these rewards do not depend on the strategies of the other player. Nash equilibrium of such stochastic games can be computed from optimal solutions of linear programs. Altman, et al., [1] considered the zero sum constrained stochastic game with discounted cost criterion where both the players have constraints. The player who controls the transition probabilities has constraints on his expected discounted costs as similar in [18], [6] and other player has subscription based constraints. This class of games also can be solved by linear programs [1].

Apart from the centralized stochastic games as discussed above, some decentralized stochastic games are being considered in the literature recently [7], [2], [19]. In decentralized stochastic games each player independently controls his own Markov chain based on his state and actions. In [2], a N -player decentralized constrained stochastic game with average cost criterion is considered and it is shown that the Nash equilibrium for these games exists in stationary strategies under the irreducibility and strong Slater condition. In these games each player controls his own Markov chain and the constraints of each player depend also on the strategies of all the players. The application of these games to modeling of wireless network is described in [7], [2], [19]. Two player zero sum game of this class where the constraints of each player do not depend on the other player's strategies is considered in [7]. These games, with both unichain and multichain structure on the state processes of both the players, can be solved by linear programs.

In this paper we consider two different classes of constrained stochastic games. First, we consider a special class of two player nonzero sum centralized constrained stochastic games which is a single controller constrained stochastic game with both average and discounted cost criterion. We then consider a N -player nonzero sum constrained stochastic game with independent state processes where all the players use average cost criterion as discussed in [2]. The summary of our results are:

1. We consider a two player nonzero sum single controller constrained stochastic game with both average and discounted cost criterion, a special class of centralized constrained stochastic games, with the same type of constraints as in [1], i.e., player 1 has subscription based constraints and player 2, who controls the transition probabilities, has realization based constraints. Unlike the situation in [1] and [6] we consider the case where realization based

constraints of player 2 depend on the strategies of both the players. It follows from [3] that there exists a stationary Nash equilibrium under strong Slater condition (irreducibility assumption is also needed in average case). We show that the Nash equilibria of this constrained stochastic game can be obtained from the global minima of one mathematical program. The converse statement is also true, i.e., from the stationary Nash equilibrium of these games we can construct a point which is a global minimum of the corresponding mathematical program.

2. If the constraints of player 2 do not depend on the strategies of player 1, then the mathematical program reduces to the non-convex quadratic program. For zero sum case the linear programs given in [1], [6] can be recovered from our quadratic program.
3. We show that the stationary Nash equilibria of N -player nonzero sum constrained stochastic game with independent state processes [2] can be obtained from the global minima of a certain mathematical program. The converse statement is also true, i.e., the stationary Nash equilibrium of these games, which exists under strong Slater and irreducibility conditions [2], corresponds to a point which is a global minimum of the corresponding mathematical program.
4. In two player constrained stochastic game with independent state processes case, if the constraints of each player do not depend on the other player's strategies, then the corresponding mathematical program reduces to the non-convex quadratic program. The linear program as given in [7] for zero sum game can be obtained as a special case of our quadratic program.

To derive mathematical programs for both constrained stochastic games we use the same approach, which is via best response linear programs. We use the fact that the best response of each player against the fixed strategies of other players can be obtained by solving a constrained Markov decision model, which, in turn, can be obtained by solving a linear program [20]. In both the cases due to some special structure we are able to put all primal-dual pair of linear programs (one pair for each player) together to form one mathematical program whose objective function is nonnegative at all feasible points. As the linear program which gives the optimal strategy in a constrained Markov decision model is given in terms of occupation measure, our mathematical programs are in terms of these occupation measures. The Nash equilibrium strategy can be recovered from occupation measure by a known transformation [20].

There are some methods available for solving non-convex quadratic programming problem [4], [21], [22]. The algorithm given in [22] is based on complete enumeration of the faces of the polyhedron and therefore it is not very efficient while the cutting plane method of [21] seems to be problematic [23]. The algorithm given in [4] to solve quadratic programs terminates in a finite number of steps. We note that the algorithm of [4] assumes that quadratic program has a global minimum and this condition is satisfied in our settings. In [5], one more algorithm based on linear programming with complementarity constraints approach is given to solve a non-convex quadratic program. This

algorithm does not assume the quadratic program to be bounded below on feasible set. (If quadratic program is not bounded below, then the algorithm given in [5] computes a feasible ray on which the quadratic program is unbounded; otherwise, it finds an optimal solution in finite number of steps). But, in our case the quadratic programs are bounded below on feasible set and hence the algorithm given in [4] is applicable to our settings. One can also attempt to use general purpose nonlinear solvers to solve these non-convex quadratic programs, but convergence to global minima may not be guaranteed.

We now describe the structure of the rest of our paper. Section 2 contains the two player nonzero sum single controller constrained stochastic game with both average and discounted cost criteria and its mathematical programming formulation. Section 3 contains N -player constrained stochastic game with independent state processes with average cost criterion and its mathematical programming formulation.

2 Single controller constrained stochastic game

We consider two player nonzero sum single controller constrained stochastic games with both average and discounted cost criterion. We assume that player 2 controls the Markov chain. As similar to [1], player 1 has subscription based constraints and player 2 has realization based constraints but unlike the case in [1], [6] the constraints of player 2 can also depend on the strategies of player 1. This class of stochastic game is described by the following objects:

- (i) S is finite state space of the game. The generic element of S is denoted by s .
- (ii) $\gamma = (\gamma(1), \gamma(2), \dots, \gamma(|S|))$ is a probability distribution over S according to which initial state is chosen.
- (iii) A^i is the finite action set of player i , $i = 1, 2$, let $A^i(s)$ denotes the set of actions available to player i when the state is at s , where $A^i = \bigcup_{s \in S} A^i(s)$.
- (iv) Define, $\mathcal{K} = \{(s, a^1, a^2) : s \in S, a^1 \in A^1(s), a^2 \in A^2(s)\}$ and for $i = 1, 2$, $\mathcal{K}^i = \{(s, a^i) : s \in S, a^i \in A^i(s)\}$.
- (v) $c^i : \mathcal{K} \rightarrow \mathbb{R}$ is immediate cost of player i , $i = 1, 2$. Specifically, $c^i(s, a^1, a^2)$ is the immediate cost incurred by player i , $i = 1, 2$, when state is $s \in S$ and actions chosen by player 1 and player 2 are $a^1 \in A^1(s)$ and $a^2 \in A^2(s)$ respectively. Player i wants to minimize the expected cost involving $c^i(\cdot)$, $i = 1, 2$.
- (vi) $d_{sub}^{1,k} : \mathcal{K}^1 \rightarrow \mathbb{R}$ is subscription type cost of player 1. $d_{sub}^{1,k}(s, a^1)$ denotes subscription cost which player 1 has to pay for using action a^1 at state s for k th service, $k = 1, 2, \dots, n_1$.
- (vii) $d^{2,l} : \mathcal{K} \rightarrow \mathbb{R}$ is immediate cost of player 2. These $d^{2,l}(\cdot)$ are involved in the l th, $l = 1, 2, \dots, n_2$, constraint on expected cost of player 2.
- (viii) Define, $\wp(M)$ as set of all probability measures over set M . $p : \mathcal{K}^2 \rightarrow \wp(S)$ is transition probability describing the dynamics of the game, where $p(s'|s, a^2)$ is a probability of going to state s' from state s when player 2 chooses action $a^2 \in A^2(s)$. We recall that the game is controlled by only player 2.

- (ix) $\xi^1 = (\xi_1^1, \xi_2^1, \dots, \xi_{n_1}^1)^T$, $\xi^2 = (\xi_1^2, \xi_2^2, \dots, \xi_{n_2}^2)^T$ denote the vectors defining the given bounds of the constraints on both the players.

The game dynamics are as follows. Initially, at time $t = 0$, the state of the game is s which is chosen according to initial distribution γ , player 1 chooses an action $a^1 \in A^1(s)$ and player 2 chooses an action $a^2 \in A^2(s)$ independent of each other. Player 1 receives an immediate cost of $c^1(s, a^1, a^2)$ and player 2 receives $c^2(s, a^1, a^2)$. Apart from this player 2 receives another immediate costs $\{d^{2,l}(s, a^1, a^2)\}$, $l = 1, 2, \dots, n_2$, which are involved in the expected cost functionals of player 2 that are constrained by specified bounds $\{\xi_l^2\}$, $l = 1, 2, \dots, n_2$. Now, the state of the game switches to a new state $\hat{s} \in S$ at time $t = 1$ with probability $p(\hat{s}|s, a^2)$. At time $t = 1$, in state \hat{s} , player 1 chooses an action $\hat{a}^1 \in A^1(\hat{s})$ and player 2 chooses an action $\hat{a}^2 \in A^2(\hat{s})$ and receives immediate cost $c^1(\hat{s}, \hat{a}^1, \hat{a}^2)$ and $c^2(\hat{s}, \hat{a}^1, \hat{a}^2)$ respectively, player 2 also receives immediate costs $\{d^{2,l}(\hat{s}, \hat{a}^1, \hat{a}^2)\}$, $l = 1, 2, \dots, n_2$. The next state of the game is $\tilde{s} \in S$ with probability $p(\tilde{s}|\hat{s}, \hat{a}^2)$. The same thing repeats at state \tilde{s} and play continues for infinite time horizon.

While transition probabilities depend only on the present state and action used, action that are used can depend on ‘past’, as in history dependent strategies. Define a history at time t as $h_t = (s_0, a_0^1, a_0^2, s_1, a_1^1, a_1^2, \dots, s_{t-1}, a_{t-1}^1, a_{t-1}^2, s_t)$ where $s_t \in S$, $a_t^i \in A^i(s_t)$, $i = 1, 2$, $t = 0, 1, 2, \dots$. Let H_t denote the set of all possible histories of length t . A decision rule $f_t : H_t \rightarrow \wp(A^1(s_t))$ (resp., $g_t : H_t \rightarrow \wp(A^2(s_t))$) of player 1 (resp., player 2) at time t is a function that assigns to any history of length t , a probability measure over action set of player 1 (resp., player 2). This means that under decision rule f_t (resp., g_t), player 1 (resp., player 2) chooses action a^1 (resp., a^2) with probability $f_t(h_t, a^1)$ (resp., $g_t(h_t, a^2)$). The sequence of decision rules is called the strategy of the player. $f^h = (f_0, f_1, \dots, f_t, \dots)$ and $g^h = (g_0, g_1, \dots, g_t, \dots)$ denote the strategies of player 1 and player 2 respectively and are called history dependent (behavioral) strategies.

Let F and G denote the set of all history dependent strategies of player 1 and player 2 respectively. These strategies are called Markovian if at every decision epoch the decision rule depends only on the current state but the decision rule can differ at every epoch. A stationary strategy is a Markovian strategy which does not depend on the time, i.e., at every decision epoch the decision rule is same. So, for stationary strategy $f_t = f$ and $g_t = g$ for all t , i.e., (f, f, f, \dots) and (g, g, g, \dots) are the stationary strategies of player 1 and player 2 respectively. We denote, with some abuse of notations, f and g as stationary strategies of player 1 and player 2 respectively. Let F_S and G_S denote the set of all stationary strategies of player 1 and player 2 respectively. A stationary strategy $f \in F_S$ is identified with $f = ((f(1))^T, (f(2))^T, \dots, (f(|S|))^T)^T$, where $f(s) = (f(s, 1), f(s, 2), \dots, f(s, |A^1(s)|))^T$ for all $s \in S$; $|M|$ denotes the cardinality of set M . Similarly, g is identified with $g = ((g(1))^T, (g(2))^T, \dots, (g(|S|))^T)^T$, where $g(s) = (g(s, 1), g(s, 2), \dots, g(s, |A^2(s)|))^T$ for all $s \in S$. For all $s \in S$, $f(s, a^1)$ is then the probability of choosing action $a^1 \in A^1(s)$ by

player 1 and $g(s, a^2)$ is probability of choosing action $a^2 \in A^2(s)$ by player 2 when state is s .

This leads to the introduction of vector stochastic process $\{X_t, \mathbb{A}_t^1, \mathbb{A}_t^2\}_{t=0}^\infty$, where, X_t denotes the state of the game, \mathbb{A}_t^1 , the action chosen by player 1 and \mathbb{A}_t^2 , the action chosen by the player 2 at time t . An initial distribution γ together with strategy pair $(f^h, g^h) \in F \times G$ defines a unique probability measure $\mathbb{P}_{f^h, g^h}^\gamma$ on an appropriate probability space with respect to which the laws of vector stochastic process $\{X_t, \mathbb{A}_t^1, \mathbb{A}_t^2\}_{t=0}^\infty$ of states and actions can be defined. The corresponding expectation operator on this probability space is denoted by $\mathbb{E}_{f^h, g^h}^\gamma$.

The expected average cost

These costs are average functionals of states and actions of the game and each player minimizes his cost functionals. For given initial distribution γ and strategy pair (f^h, g^h) the expected average cost of player i , $i = 1, 2$, is defined as

$$C_{ea}^i(\gamma, f^h, g^h) = \limsup_{T \rightarrow \infty} \frac{1}{T} \sum_{t=0}^{T-1} \mathbb{E}_{f^h, g^h}^\gamma c^i(X_t, \mathbb{A}_t^1, \mathbb{A}_t^2) \quad (1)$$

where ea stands for expected average.

The expected average constraints

The expected average constraints of player 2 are defined by average functionals of states and actions of the game which are bounded by given reals. For given initial distribution γ and strategy pair (f^h, g^h) the expected average costs of player 2 are defined as

$$D_{ea}^{2,l}(\gamma, f^h, g^h) = \limsup_{T \rightarrow \infty} \frac{1}{T} \sum_{t=0}^{T-1} \mathbb{E}_{f^h, g^h}^\gamma d^{2,l}(X_t, \mathbb{A}_t^1, \mathbb{A}_t^2), \quad \forall l = 1, 2, \dots, n_2.$$

$D_{ea}^{2,l}(\cdot, \cdot)$ can capture the average consumption of resource l , $l = 1, 2, \dots, n_2$, by player 2. The expected average constraints of player 2 are given by

$$D_{ea}^{2,l}(\gamma, f^h, g^h) \leq \xi_l^2, \quad \forall l = 1, 2, \dots, n_2. \quad (2)$$

A constraint in (2) captures the fact that the average consumption of resource l by player 2, when player 1 uses strategy f^h and player 2 uses strategy g^h is not more than given constant ξ_l^2 , $l = 1, 2, \dots, n_2$.

The expected discounted cost

These costs are discounted functionals of states and actions of the game and each player minimizes his cost functionals. For given initial distribution γ and strategy pair (f^h, g^h) the expected discounted cost of player i , $i = 1, 2$, is defined as

$$C_\beta^i(\gamma, f^h, g^h) = (1 - \beta) \sum_{t=0}^{\infty} \beta^t \mathbb{E}_{f^h, g^h}^\gamma c^i(X_t, \mathbb{A}_t^1, \mathbb{A}_t^2) \quad (3)$$

where $\beta \in [0, 1)$ is a fixed discount factor.

The expected discounted constraints

The expected discounted constraints of player 2 are defined by discounted functionals of states and actions of the game which are bounded by given reals. For given initial distribution γ and strategy pair (f^h, g^h) the expected discounted costs of player 2 are defined as

$$D_\beta^{2,l}(\gamma, f^h, g^h) = (1 - \beta) \sum_{t=0}^{\infty} \beta^t \mathbb{E}_{f^h, g^h}^\gamma d^{2,l}(X_t, \mathbb{A}_t^1, \mathbb{A}_t^2), \quad \forall l = 1, 2, \dots, n_2.$$

$D_\beta^{2,l}(\cdot, \cdot)$ can capture the discounted cost for the consumption of resource l , $l = 1, 2, \dots, n_2$, by player 2. The expected discounted constraints of player 2 are given by

$$D_\beta^{2,l}(\gamma, f^h, g^h) \leq \xi_l^2, \quad \forall l = 1, 2, \dots, n_2. \quad (4)$$

A constraint in (4) captures the fact that discounted cost for the consumption of resource l by player 2, when player 1 uses strategy f^h and player 2 uses strategy g^h is not more than given real ξ_l^2 , $l = 1, 2, \dots, n_2$.

Subscription type cost [1]

The subscription type costs of player 1 are defined as in [1]

$$D_{sub}^{1,k}(f) = \sum_{s \in S} \sum_{a^1 \in A^1(s)} d_{sub}^{1,k}(s, a^1) f(s, a^1)$$

for all $k = 1, 2, \dots, n_1$ and $f \in F_S$. These costs are called as subscription type because they are based only on the fraction of time during which a given action is used at a given state and are not based on how frequently the state is visited and action is used. This situation can arise where for using some services there is subscription/registration fee for their planned use and that can be paid in advance.

Subscription type constraints

The subscription type constraints of player 1 are defined as

$$D_{sub}^{1,k}(f) \leq \xi_k^1, \quad \forall k = 1, 2, \dots, n_1. \quad (5)$$

We denote C^i , $i = 1, 2$ and $D^{2,l}$, $l = 1, 2, \dots, n_2$, as expected costs which can be either average or discounted that depends on the criterion being used. Under average cost criterion both players have expected average costs and under discounted cost criterion both players have expected discounted costs. Apart from this, player 1 has subscription based costs which are constrained by some given reals. It is clear that player 1 has n_1 number of constraints which are defined by (5) and player 2 has n_2 number of constraints which are defined as

$$D^{2,l}(\gamma, f^h, g^h) \leq \xi_l^2, \quad \forall l = 1, 2, \dots, n_2. \quad (6)$$

The constraints (5) and (6) are called subscription based and realization based constraints respectively. Both the players choose their actions independently and want to minimize their expected cost subject to their constraints from (5) and (6). We denote this constrained stochastic game by G^c . As Nash equilibrium exists in stationary strategies under assumptions (A1)-(A2) given below [3], from now onwards we restrict ourselves to the stationary strategies.

The strategy pair (f, g) is called 1-feasible if it satisfies (5) and strategy pair (f, g) is called 2-feasible if it satisfies (6). As the player 1 constraints (5) do not depend on the strategies of player 2, then strategy pair (f, g) is 1-feasible for all $g \in G_S$ if f satisfies (5). A strategy pair (f, g) is called feasible if it is both 1-feasible and 2-feasible. Let F_S^ξ denote the set of all feasible stationary strategy pairs for the constrained stochastic game G^c . We shall assume throughout that F_S^ξ is non-empty. Now, we recall the definition of Nash equilibrium as given in [3]. A strategy pair $(f^*, g^*) \in F_S^\xi$ is called the Nash equilibrium of constrained stochastic game G^c if it satisfies the following conditions

$$C^1(\gamma, f^*, g^*) \leq C^1(\gamma, f, g^*), \quad \forall \text{ 1-feasible } (f, g^*) \quad (7)$$

$$C^2(\gamma, f^*, g^*) \leq C^2(\gamma, f^*, g^h), \quad \forall \text{ 2-feasible } (f^*, g^h). \quad (8)$$

Thus, unilateral deviation of any player i , $i = 1, 2$, will either violate the constraints of i th player, or if it does not, it will result in a cost C^i for that player that is not lower than the one achieved by feasible strategy pair (f^*, g^*) . The strategy pair $(f^*, g^*) \in F_S^\xi$ satisfying (7) and (8) would still be Nash equilibrium of constrained stochastic game if we replace strategy g^h by stationary strategy g in (8). This can be seen by noticing that when strategy of player 1 is fixed as a stationary strategy f^* , then player 2 is faced with a constrained Markov decision process (CMDP) where optimal strategy always exists in the space of stationary strategies [20].

Assumptions [Altman and Shwartz [3]]

- (A1) Ergodicity: In case of average cost criterion the unichain ergodic structure holds, i.e., under every stationary strategy g the state process is an irreducible Markov chain with one ergodic class (and possibly some transient states).
- (A2) Strong Slater condition: For player 2, there exists some g' such that for any strategy f of player 1,

$$D^{2,l}(\gamma, f, g') < \xi_l^2, \quad \forall l = 1, 2, \dots, n_2.$$

As the constraints of player 1 are linear and does not depend on the strategies of player 2, the strong Slater condition is not needed for the constraints of player 1.

We use the following notations throughout this section. For $i = 1, 2, s \in S, l = 1, 2, \dots, n_2$,

- $C^i(s) = [c^i(s, a^1, a^2)]_{a^1=1, a^2=1}^{|A^1(s)|, |A^2(s)|}$.
- $C^i = \text{diag}(C^i(1), C^i(2), \dots, C^i(|S|))$.
- $D^{2,l}(s) = [d^{2,l}(s, a^1, a^2)]_{a^1=1, a^2=1}^{|A^1(s)|, |A^2(s)|}$.
- $x = (x(1)^T, x(2)^T, \dots, x(|S|)^T)^T$.
- $x(s) = (x(s, 1), x(s, 2), \dots, x(s, |A^2(s)|))^T$.
- $u = (u(1), u(2), \dots, u(|S|))^T$.
- $v \in \mathbb{R}$.
- $z = (z(1), z(2), \dots, z(|S|))^T$.
- $\delta^1 = (\delta_1^1, \delta_2^1, \dots, \delta_{n_1}^1)^T$.
- $\delta^2 = (\delta_1^2, \delta_2^2, \dots, \delta_{n_2}^2)^T$.
- $\mathbf{1}_n = (1, 1, \dots, 1)^T \in \mathbb{R}^n$.

2.1 Single controller constrained stochastic game with average cost criterion

In this section we consider the game described in Section 2 with average cost criterion where both players choose their strategies independently and minimize their expected average costs as defined in (1) subject to their constraints from (5), (2). The constraints of player 1 given in (5) are subscription based. The expected average constraints (2) of player 2 captures the fact that the average consumption of resource $l, l = 1, 2, \dots, n_2$, by player 2 is not more than given ξ_l^2 .

2.1.1 Average occupation measure

For an initial distribution γ and a stationary strategy g define the average occupation measure

$$\pi_{ea}^2(\gamma, g) := \{\pi_{ea}^2(\gamma, g; s, a^2) : s \in S, a^2 \in A^2(s)\}.$$

For all $s \in S$, $a^2 \in A^2(s)$, $\pi_{ea}^2(\gamma, g; s, a^2)$ is given by

$$\pi_{ea}^2(\gamma, g; s, a^2) = \pi^g(s)g(s, a^2) \quad (9)$$

where $\pi^g = (\pi^g(1), \pi^g(2), \dots, \pi^g(|S|))$ is steady state distribution of Markov chain induced by stationary strategy g which exists and is unique under (A1). $\pi_{ea}^2(\gamma, g)$ can be considered as a probability measure over \mathcal{K}^2 that assigns probability $\pi_{ea}^2(\gamma, g; s, a^2)$ to the state-action pair (s, a^2) . The occupation measure defined as in (9) is independent from initial distribution γ , so, we drop γ from the notation. For fixed strategy pair $(f, g) \in F_S \times G_S$ the expected average costs of both the players can be written in terms of occupation measure as

$$C_{ea}^i(f, g) = \sum_{(s, a^2) \in \mathcal{K}^2} \pi_{ea}^2(g; s, a^2) \sum_{a^1 \in A^1(s)} f(s, a^1) c^i(s, a^1, a^2), \quad \forall i = 1, 2.$$

$$D_{ea}^{2,l}(f, g) = \sum_{(s, a^2) \in \mathcal{K}^2} \pi_{ea}^2(g; s, a^2) \sum_{a^1 \in A^1(s)} f(s, a^1) d^{2,l}(s, a^1, a^2)$$

for all $l = 1, 2, \dots, n_2$.

Let Q_{ea} be the set of vectors $x \in \mathbb{R}^{|\mathcal{K}^2|}$ satisfying

$$\begin{cases} (i) & \sum_{(s, a^2) \in \mathcal{K}^2} (\delta(s, s') - p(s'|s, a^2)) x(s, a^2) = 0, \quad \forall s' \in S \\ (ii) & \sum_{(s, a^2) \in \mathcal{K}^2} x(s, a^2) = 1 \\ (iii) & x(s, a^2) \geq 0, \quad \forall s \in S, a^2 \in A^2(s). \end{cases}$$

$\delta(\cdot, \cdot)$ is a Kronecker delta, i.e.,

$$\delta(s, s') = \begin{cases} 1 & \text{if } s = s', \\ 0 & \text{if } s \neq s'. \end{cases}$$

The stationary strategies are complete, i.e., set of occupation measures achieved by history dependent strategies equals to those achieved by stationary strategies and further equals to the set Q_{ea} [20]. It is known that for each $(s, a^2) \in \mathcal{K}^2$, $x(s, a^2) = \pi_{ea}^2(g; s, a^2)$ where

$$g(s, a^2) = \frac{x(s, a^2)}{\sum_{a^2 \in A^2(s)} x(s, a^2)} \quad (10)$$

whenever denominator is nonzero (when it is zero $g(s)$ is chosen arbitrarily from $\wp(A^2(s))$) [20].

The cost of player 1 when he uses action a^1 at state s and player 2 uses strategy g is given by

$$c^1(s, a^1; g) = \sum_{(s, a^2) \in \mathcal{K}^2} c^1(s, a^1, a^2) \pi_{ea}^2(g; s, a^2).$$

Similarly, the costs of player 2 when he uses action a^2 at state s and player 1 uses strategy f are given by

$$c^2(f; s, a^2) = \sum_{a^1 \in A^1(s)} c^2(s, a^1, a^2) f(s, a^1).$$

$$d^{2,l}(f; s, a^2) = \sum_{a^1 \in A^1(s)} d^{2,l}(s, a^1, a^2) f(s, a^1), \quad \forall l = 1, 2, \dots, n_2.$$

2.1.2 Mathematical programming formulation

We show the one to one correspondence between the stationary Nash equilibria of single controller constrained stochastic game G^c with average cost criterion and the global minima of a certain mathematical program.

Best response linear programs

For a given stationary strategy of one player in a two player constrained stochastic game, the best response of the other player is given by solving a constrained Markov decision model, which, in turn, can be obtained by a linear program in finite state-action setting [20]. For fixed strategy g of player 2, the best response of player 1 can be obtained from the following linear program:

$$\left. \begin{array}{l} \min_f \sum_{(s, a^1) \in \mathcal{K}^1} c^1(s, a^1; g) f(s, a^1) \\ \text{s.t.} \\ (i) \quad \sum_{(s, a^1) \in \mathcal{K}^1} d_{sub}^{1,k}(s, a^1) f(s, a^1) \leq \xi_k^1, \quad \forall k = 1, 2, \dots, n_1 \\ (ii) \quad \sum_{a^1 \in A^1(s)} f(s, a^1) = 1, \quad \forall s \in S \\ (iii) \quad f(s, a^1) \geq 0, \quad \forall s \in S, a^1 \in A^1(s). \end{array} \right\} \quad (11)$$

The dual of (11) is

$$\left. \begin{array}{l} \max_{z, \delta^1} \left[\sum_{s \in S} z(s) - \sum_{k=1}^{n_1} \delta_k^1 \xi_k^1 \right] \\ \text{s.t.} \\ (i) \quad z(s) \leq c^1(s, a^1; g) + \sum_{k=1}^{n_1} \delta_k^1 d_{sub}^{1,k}(s, a^1), \quad \forall s \in S, a^1 \in A^1(s) \\ (ii) \quad \delta_k^1 \geq 0, \quad \forall k = 1, 2, \dots, n_1. \end{array} \right\} \quad (12)$$

Similarly, for fixed strategy f of player 1, the best response of player 2 can be obtained from the following linear program:

$$\left. \begin{array}{l} \min_x \sum_{(s,a^2) \in \mathcal{K}^2} c^2(f; s, a^2) x(s, a^2) \\ \text{s.t.} \\ (i) \quad \sum_{(s,a^2) \in \mathcal{K}^2} (\delta(s, s') - p(s'|s, a^2)) x(s, a^2) = 0, \quad \forall s' \in S \\ (ii) \quad \sum_{(s,a^2) \in \mathcal{K}^2} x(s, a^2) = 1 \\ (iii) \quad \sum_{(s,a^2) \in \mathcal{K}^2} d^{2,l}(f; s, a^2) x(s, a^2) \leq \xi_l^2, \quad \forall l = 1, 2, \dots, n_2 \\ (iv) \quad x(s, a^2) \geq 0, \quad \forall s \in S, a^2 \in A^2(s). \end{array} \right\} \quad (13)$$

If x^* is the optimal solution of the linear program (13), then the best response strategy g^* of player 2 can be obtained from (10) [20]. The dual of the linear program (13) is given by

$$\left. \begin{array}{l} \max_{v, u, \delta^2} \left[v - \sum_{l=1}^{n_2} \delta_l^2 \xi_l^2 \right] \\ \text{s.t.} \\ (i) \quad v + u(s) \leq c^2(f; s, a^2) + \sum_{l=1}^{n_2} \delta_l^2 d^{2,l}(f; s, a^2) \\ \quad \quad \quad + \sum_{s' \in S} p(s'|s, a^2) u(s'), \quad \forall s \in S, a^2 \in A^2(s) \\ (ii) \quad \delta_l^2 \geq 0, \quad \forall l = 1, 2, \dots, n_2. \end{array} \right\} \quad (14)$$

We denote the decision variables and objective function of mathematical program [MP1] by $\eta = (v, u^T, z^T, f^T, x^T, (\delta^1)^T, (\delta^2)^T)^T$ and $\Phi(\eta)$ respectively.

Theorem 1 (a) *If (f^*, g^*) is a Nash equilibrium of the constrained stochastic game G^c with average cost criterion, then, there exists a vector $\eta^* = (v^*, u^{*T}, z^{*T}, f^{*T}, x^{*T}, (\delta^{1*})^T, (\delta^{2*})^T)^T$ such that it is a global minimum of mathematical program [MP1] given below*

$$\begin{aligned} [\text{MP1}] \quad & \min_{\eta} \left[(f^T C^1 x - (I_{|S|}^T z - (\delta^1)^T \xi^1)) + (f^T C^2 x - (v - (\delta^2)^T \xi^2)) \right] \\ & \text{s.t.} \\ (i) \quad & v + u(s) \leq \left[(f(s))^T C^2(s) \right]_{a^2} + \sum_{l=1}^{n_2} \delta_l^2 \left[(f(s))^T D^{2,l}(s) \right]_{a^2} \\ & \quad + \sum_{s' \in S} p(s'|s, a^2) u(s'), \quad \forall s \in S, a^2 \in A^2(s) \end{aligned}$$

- (ii) $z(s) \leq [C^1(s)x(s)]_{a^1} + \sum_{k=1}^{n_1} \delta_k^1 d_{sub}^{1,k}(s, a^1), \quad \forall s \in S, a^1 \in A^1(s)$
- (iii) $\sum_{(s, a^2) \in \mathcal{K}^2} [\delta(s, s') - p(s'|s, a^2)] x(s, a^2) = 0, \quad \forall s' \in S$
- (iv) $\sum_{(s, a^2) \in \mathcal{K}^2} x(s, a^2) = 1$
- (v) $\sum_{(s, a^1) \in \mathcal{K}^1} d_{sub}^{1,k}(s, a^1) f(s, a^1) \leq \xi_k^1, \quad \forall k = 1, 2, \dots, n_1$
- (vi) $\sum_{s \in S} (f(s))^T D^{2,l}(s) x(s) \leq \xi_l^2, \quad \forall l = 1, 2, \dots, n_2$
- (vii) $\sum_{a^1 \in A^1(s)} f(s, a^1) = 1, \quad \forall s \in S$
- (viii) $f(s, a^1) \geq 0, \quad \forall s \in S, a^1 \in A^1(s)$
- (ix) $x(s, a^2) \geq 0, \quad \forall s \in S, a^2 \in A^2(s)$
- (x) $\delta_k^1 \geq 0, \quad \forall k = 1, 2, \dots, n_1$
- (xi) $\delta_l^2 \geq 0, \quad \forall l = 1, 2, \dots, n_2.$

with $\Phi(\eta^*) = 0$.

- (b) If $\eta^* = (v^*, u^{*T}, z^{*T}, f^{*T}, x^{*T}, (\delta^{1*})^T, (\delta^{2*})^T)^T$ is a global minimum of [MP1] with $\Phi(\eta^*) = 0$, then, (f^*, g^*) is a Nash equilibrium of the constrained stochastic game G^c with average cost criterion, where

$$g^*(s, a^2) = \frac{x^*(s, a^2)}{\sum_{a^2 \in A^2(s)} x^*(s, a^2)}$$

for all $s \in S, a^2 \in A^2(s)$ whenever the denominator is non-zero (when it is zero $g^*(s)$ is chosen arbitrarily from $\wp(A^2(s))$).

Proof (a) Let (f^*, g^*) be Nash equilibrium of the constrained stochastic game G^c with average cost criterion. We construct occupation measure x^* corresponding to g^* as given in (9) then x^* satisfies (iii), (iv) and (ix) of [MP1]. The strategy pair (f^*, g^*) is feasible because it is a Nash equilibrium, so, (f^*, x^*) satisfy (v)-(viii) of [MP1]. As f^* and g^* are best responses of each other, x^* as constructed above will be optimal solution of linear program (13) for fixed f^* from Proposition 3.1(ii) of [3]. By strong duality theorem [24], [25] there exists optimal solution (v^*, u^*, δ^{2*}) of (14) such that $(v^*, u^*, f^*, \delta^{2*})$ satisfy (i) and (xi) of [MP1] and objective function value of (13) and (14) are equal. Similarly, f^* is an optimal solution of linear program (11) for fixed g^* and hence there exists optimal solution (z^*, δ^{1*}) of (12) such that (z^*, x^*, δ^{1*}) satisfy (ii) and (x) of [MP1] and objective function value of (11) and (12) are equal. In other words we have a point $\eta^* = (v^*, u^{*T}, z^{*T}, f^{*T}, x^{*T}, (\delta^{1*})^T, (\delta^{2*})^T)^T$ such that (i), (ii), (x) and (xi) are satisfied and

$$f^{*T} C^1 x^* = \mathbf{1}_{|S|}^T z^* - (\delta^{1*})^T \xi^1,$$

$$f^{*T} \mathbf{C}^2 x^* = v^* - (\delta^{2*})^T \xi^2.$$

Thus, η^* is a feasible point of the mathematical program [MP1] and from the construction of the objective function, $\Phi(\eta^*) = 0$.

Let η be any feasible point of [MP1]. Multiply each constraint in (ii) of [MP1] corresponding to pair (s, a^1) by $f(s, a^1)$ and then add over all $(s, a^1) \in \mathcal{K}^1$ and by using the constraints (v), (vii), (viii) and (x) we have

$$f^T \mathbf{C}^1 x \geq \mathbf{1}_{|S|}^T z - (\delta^1)^T \xi^1. \quad (15)$$

By using the similar arguments as above, i.e., multiply each constraint in (i) of [MP1] corresponding to pair (s, a^2) by $x(s, a^2)$ and add over all $(s, a^2) \in \mathcal{K}^2$ and by using the constraints (iii), (iv), (vi), (ix) and (xi), we have

$$f^T \mathbf{C}^2 x \geq v - (\delta^2)^T \xi^2. \quad (16)$$

We have from (15) and (16), $\Phi(\eta) \geq 0$ for all feasible points η of [MP1]. Thus η^* is a global minimum of the [MP1].

(b) Let η^* be a global minimum of [MP1] such that $\Phi(\eta^*) = 0$. As η^* is a feasible point of [MP1] then (15) and (16) will also hold for η^* , i.e.,

$$\begin{aligned} f^{*T} \mathbf{C}^1 x^* &\geq \mathbf{1}_{|S|}^T z^* - (\delta^{1*})^T \xi^1 \\ f^{*T} \mathbf{C}^2 x^* &\geq v^* - (\delta^{2*})^T \xi^2. \end{aligned}$$

From above, both the terms of objective function are non-negative at η^* but the objective function value is zero at η^* which means both the terms are individually zero, i.e.,

$$\left. \begin{aligned} f^{*T} \mathbf{C}^1 x^* &= \mathbf{1}_{|S|}^T z^* - (\delta^{1*})^T \xi^1 \\ f^{*T} \mathbf{C}^2 x^* &= v^* - (\delta^{2*})^T \xi^2. \end{aligned} \right\} \quad (17)$$

Fix η^* , and from the same argument used as in (15) and by using the constraints (v), (vii), (viii), (x) and (17) we have the following inequality

$$f^{*T} \mathbf{C}^1 x^* \leq f^T \mathbf{C}^1 x^*, \quad \forall \text{ 1-feasible } (f, x^*),$$

Similarly we have

$$f^{*T} \mathbf{C}^2 x^* \leq f^{*T} \mathbf{C}^2 x, \quad \forall \text{ 2-feasible } (f^*, x)$$

In other words we can say that

$$\begin{aligned} C_{ea}^1(f^*, g^*) &\leq C_{ea}^1(f, g^*), \quad \forall \text{ 1-feasible } (f, g^*) \\ C_{ea}^2(f^*, g^*) &\leq C_{ea}^2(f^*, g), \quad \forall \text{ 2-feasible } (f^*, g), \end{aligned}$$

where

$$g^*(s, a^2) = \frac{x^*(s, a^2)}{\sum_{a^2 \in A^2(s)} x^*(s, a^2)}$$

for all $s \in S$, $a^2 \in A^2(s)$ whenever the denominator is non-zero (when it is zero $g^*(s)$ is chosen arbitrarily from $\varnothing(A^2(s))$). This implies that (f^*, g^*) is a Nash equilibrium of the constrained stochastic game G^c with average cost criterion.

Remark 1 Because the diagonal elements of the objective function's Hessian matrix are zero, it will have some positive as well as some negative eigenvalues. So, the objective function of [MP1] is a non-convex function. As there are some non-convex constraints, the feasible region is also not a convex set. So, [MP1] is a non-convex constrained optimization problem.

2.1.3 Special cases

We consider two special cases. First, we consider nonzero sum game as defined in Section 2 with average cost criterion where the constraints of player 2 do not depend on the strategies of player 1. Next, we briefly consider the zero sum game as considered in [6].

(i) Quadratic program in the case of decoupled constraints

We consider the situation where the constraints of player 2 do not depend on the strategies of the player 1. This is possible when the immediate costs of player 2 which correspond to the constraints of player 2 do not depend on the actions of player 1, i.e.,

$$d^{2,l}(s, a^1, a^2) = d^{2,l}(s, a^2), \quad \forall s \in S, a^1 \in A^1(s), a^2 \in A^2(s) \text{ and } \forall l = 1, 2, \dots, n_2. \quad (18)$$

Under this condition [MP1] reduces to the quadratic program [QP1] given below:

$$[\text{QP1}] \quad \min_{\eta} \left[\left(f^T C^1 x - \left(\mathbf{1}_{|S|}^T z - (\delta^1)^T \xi^1 \right) \right) + \left(f^T C^2 x - (v - (\delta^2)^T \xi^2) \right) \right]$$

s.t.

$$\begin{aligned} (i) \quad & v + u(s) \leq [(f(s))^T C^2(s)]_{a^2} + \sum_{l=1}^{n_2} \delta_l^2 d^{2,l}(s, a^2) \\ & + \sum_{s' \in S} p(s'|s, a^2) u(s'), \quad \forall s \in S, a^2 \in A^2(s) \\ (ii) \quad & z(s) \leq [C^1(s)x(s)]_{a^1} + \sum_{k=1}^{n_1} \delta_k^1 d_{sub}^{1,k}(s, a^1), \quad \forall s \in S, a^1 \in A^1(s) \\ (iii) \quad & \sum_{(s, a^2) \in \mathcal{K}^2} [\delta(s, s') - p(s'|s, a^2)] x(s, a^2) = 0, \quad \forall s' \in S \\ (iv) \quad & \sum_{(s, a^2) \in \mathcal{K}^2} x(s, a^2) = 1 \\ (v) \quad & \sum_{(s, a^1) \in \mathcal{K}^1} d_{sub}^{1,k}(s, a^1) f(s, a^1) \leq \xi_k^1, \quad \forall k = 1, 2, \dots, n_1 \\ (vi) \quad & \sum_{(s, a^2) \in \mathcal{K}^2} d^{2,l}(s, a^2) x(s, a^2) \leq \xi_l^2, \quad \forall l = 1, 2, \dots, n_2 \end{aligned}$$

- (vii) $\sum_{a^1 \in A^1(s)} f(s, a^1) = 1, \forall s \in S$
- (viii) $f(s, a^1) \geq 0, \forall s \in S, a^1 \in A^1(s)$
- (ix) $x(s, a^2) \geq 0, \forall s \in S, a^2 \in A^2(s)$
- (x) $\delta_k^1 \geq 0, \forall k = 1, 2, \dots, n_1$
- (xi) $\delta_l^2 \geq 0, \forall l = 1, 2, \dots, n_2$.

(ii) *Zero sum single controller constrained stochastic games*

The zero sum single controller constrained stochastic game with average cost criterion is considered in [6]. We assume that player 1 minimizes the expected average cost of the game and player 2 has opposite objective, i.e., he maximizes the expected average cost of the game. In [6], the player who controls the transition probabilities has realization based constraints and other player has no constraints and these games can be solved by a linear program. By substituting $C^1(s) = -C^2(s) = C(s)$ for all $s \in S$ and without the subscription type constraints, the quadratic program [QP1] can be reduced into primal-dual pair of linear programs which are same as given in [6].

2.2 Single controller constrained stochastic game with discounted cost criterion

In this section we consider the game described in Section 2 with discounted cost criterion where both players choose their strategies independently and minimize their expected discounted costs as defined in (3) subject to their constraints from (5), (4). The constraints of player 1 given in (5) are subscription based. The expected discounted constraints (4) of player 2 captures the fact that discounted cost for the consumption of resource l , $l = 1, 2, \dots, n_2$, by player 2 is not more than given ξ_l^2 . Similar to the average cost criterion we give one mathematical program which characterizes stationary Nash equilibria of these games.

2.2.1 Discounted occupation measure

For an initial distribution γ and a stationary strategy g define the discounted occupation measure

$$\pi_\beta^2(\gamma, g) := \{ \pi_\beta^2(\gamma, g; s, a^2) : s \in S, a^2 \in A^2(s) \}.$$

For all $s \in S, a^2 \in A^2(s)$, $\pi_\beta^2(\gamma, g; s, a^2)$ is given by

$$\pi_\beta^2(\gamma, g; s, a^2) = (1 - \beta) \left(\sum_{t=0}^{\infty} \beta^t \sum_{s' \in S} \gamma(s') ([P(g)]^t)_{s's} \right) g(s, a^2), \quad (19)$$

here $[P(g)]^0$ is the identity matrix. $\pi_\beta^2(\gamma, g)$ can be considered as a probability measure over \mathcal{K}^2 that assigns probability $\pi_\beta^2(\gamma, g; s, a^2)$ to the state-action pair (s, a^2) . For fixed strategy pair $(f, g) \in F_S \times G_S$ the expected discounted costs of both players can be written in terms of occupation measure as

$$C_\beta^i(\gamma, f, g) = \sum_{(s, a^2) \in \mathcal{K}^2} \pi_\beta^2(\gamma, g; s, a^2) \sum_{a^1 \in A^1(s)} f(s, a^1) c^i(s, a^1, a^2), \quad \forall i = 1, 2.$$

$$D_\beta^{2,l}(\gamma, f, g) = \sum_{(s, a^2) \in \mathcal{K}^2} \pi_\beta^2(\gamma, g; s, a^2) \sum_{a^1 \in A^1(s)} f(s, a^1) d^{2,l}(s, a^1, a^2)$$

for all $l = 1, 2, \dots, n_2$.

Let $Q^\beta(\gamma)$ be the set of vectors $x \in \mathbb{R}^{|\mathcal{K}^2|}$ satisfying

$$\begin{cases} (i) \sum_{(s, a^2) \in \mathcal{K}^2} (\delta(s, s') - \beta p(s'|s, a^2)) x(s, a^2) = (1 - \beta) \gamma(s'), \quad \forall s' \in S \\ (ii) x(s, a^2) \geq 0, \quad \forall s \in S, a^2 \in A^2(s). \end{cases}$$

By summing the first constraint over s' we note that $\sum_{(s, a^2) \in \mathcal{K}^2} x(s, a^2) = 1$, so the x satisfying the above constraints are probability measures. The stationary strategies are complete, i.e., set of occupation measures achieved by history dependent strategies equals to those achieved by stationary strategies and further equals to the set $Q^\beta(\gamma)$ [20]. It is known that for each $(s, a^2) \in \mathcal{K}^2$, $x(s, a^2) = \pi_\beta^2(\gamma, g; s, a^2)$ where

$$g(s, a^2) = \frac{x(s, a^2)}{\sum_{a^2 \in A^2(s)} x(s, a^2)} \quad (20)$$

whenever denominator is nonzero (when it is zero $g(s)$ is chosen arbitrarily from $\wp(A^2(s))$) [20].

The cost of player 1 when he uses action a^1 at state s and player 2 uses strategy g is given by

$$c^1(s, a^1; g) = \sum_{(s, a^2) \in \mathcal{K}^2} c^1(s, a^1, a^2) \pi_\beta^2(\gamma, g; s, a^2).$$

Similarly, the costs of player 2 when he uses action a^2 at state s and player 1 uses strategy f are given by

$$\begin{aligned} c^2(f; s, a^2) &= \sum_{a^1 \in A^1(s)} c^2(s, a^1, a^2) f(s, a^1). \\ d^{2,l}(f; s, a^2) &= \sum_{a^1 \in A^1(s)} d^{2,l}(s, a^1, a^2) f(s, a^1), \quad \forall l = 1, 2, \dots, n_2. \end{aligned}$$

2.2.2 Mathematical programming formulation

Similar to average cost criterion we show the one to one correspondence between the stationary Nash equilibria of this class of game and the global minima of a certain mathematical program.

Best response linear programs

For fixed strategy g of player 2, the best response of player 1 can be obtained from the following linear program:

$$\left. \begin{array}{l} \min_f \sum_{(s,a^1) \in \mathcal{K}^1} c^1(s, a^1; g) f(s, a^1) \\ \text{s.t.} \\ (i) \quad \sum_{(s,a^1) \in \mathcal{K}^1} d_{sub}^{1,k}(s, a^1) f(s, a^1) \leq \xi_k^1, \quad \forall k = 1, 2, \dots, n_1 \\ (ii) \quad \sum_{a^1 \in A^1(s)} f(s, a^1) = 1, \quad \forall s \in S \\ (iii) \quad f(s, a^1) \geq 0, \quad \forall s \in S, a^1 \in A^1(s). \end{array} \right\} \quad (21)$$

The dual of (21) is

$$\left. \begin{array}{l} \max_{z, \delta^1} \left[\sum_{s \in S} z(s) - \sum_{k=1}^{n_1} \delta_k^1 \xi_k^1 \right] \\ \text{s.t.} \\ (i) \quad z(s) \leq c^1(s, a^1; g) + \sum_{k=1}^{n_1} \delta_k^1 d_{sub}^{1,k}(s, a^1), \quad \forall s \in S, a^1 \in A^1(s) \\ (ii) \quad \delta_k^1 \geq 0, \quad \forall k = 1, 2, \dots, n_1. \end{array} \right\} \quad (22)$$

Similarly for fixed strategy f of player 1, the best response of player 2 can be obtained from the following linear program:

$$\left. \begin{array}{l} \min_x \sum_{(s,a^2) \in \mathcal{K}^2} c^2(f; s, a^2) x(s, a^2) \\ \text{s.t.} \\ (i) \quad \sum_{(s,a^2) \in \mathcal{K}^2} (\delta(s, s') - \beta p(s'|s, a^2)) x(s, a^2) = (1 - \beta) \gamma(s'), \quad \forall s' \in S \\ (ii) \quad \sum_{(s,a^2) \in \mathcal{K}^2} d^{2,l}(f; s, a^2) x(s, a^2) \leq \xi_l^2, \quad \forall l = 1, 2, \dots, n_2 \\ (iii) \quad x(s, a^2) \geq 0, \quad \forall s \in S, a^2 \in A^2(s). \end{array} \right\} \quad (23)$$

If x^* is the optimal solution of the linear program (23) then the best response strategy g^* of player 2 can be obtained from (20) [20]. The dual of the linear program (23) is given by

$$\left. \begin{aligned} & \max_{u, \delta^2} \left[\sum_{s \in S} (1 - \beta) \gamma(s) u(s) - \sum_{l=1}^{n_2} \delta_l^2 \xi_l^2 \right] \\ & \text{s.t.} \\ & (i) \ u(s) \leq c^2(f; s, a^2) + \sum_{l=1}^{n_2} \delta_l^2 d^{2,l}(f; s, a^2) \\ & \quad + \beta \sum_{s' \in S} p(s'|s, a^2) u(s'), \quad \forall \ s \in S, \ a^2 \in A^2(s) \\ & (ii) \ \delta_l^2 \geq 0, \quad \forall \ l = 1, 2, \dots, n_2. \end{aligned} \right\} \quad (24)$$

By using the best response linear programs (21), (22), (23), (24) we have similar results as in the case of average cost criterion.

Theorem 2 (a) *If (f^*, g^*) is a Nash equilibrium of the constrained stochastic game G^c with discounted cost criterion, then, there exists a vector $\eta^* = (u^{*T}, z^{*T}, f^{*T}, x^{*T}, (\delta^{1*})^T, (\delta^{2*})^T)^T$ such that it is a global minimum of mathematical program [MP2] given below*

$$\begin{aligned} [\text{MP2}] \quad & \min_{\eta} \left[\left(f^T C^1 x - \left(\mathbf{1}_{|S|}^T z - (\delta^1)^T \xi^1 \right) \right) \right. \\ & \left. + \left(f^T C^2 x - ((1 - \beta) \gamma^T u - (\delta^2)^T \xi^2) \right) \right] \\ & \text{s.t.} \\ & (i) \ u(s) \leq [(f(s))^T C^2(s)]_{a^2} + \sum_{l=1}^{n_2} \delta_l^2 [(f(s))^T D^{2,l}(s)]_{a^2} \\ & \quad + \beta \sum_{s' \in S} p(s'|s, a^2) u(s'), \quad \forall \ s \in S, \ a^2 \in A^2(s) \\ & (ii) \ z(s) \leq [C^1(s)x(s)]_{a^1} + \sum_{k=1}^{n_1} \delta_k^1 d_{sub}^{1,k}(s, a^1), \quad \forall \ s \in S, \ a^1 \in A^1(s) \\ & (iii) \ \sum_{(s, a^2) \in \mathcal{K}^2} [\delta(s, s') - \beta p(s'|s, a^2)] x(s, a^2) = (1 - \beta) \gamma(s'), \quad \forall \ s' \in S \\ & (iv) \ \sum_{(s, a^1) \in \mathcal{K}^1} d_{sub}^{1,k}(s, a^1) f(s, a^1) \leq \xi_k^1, \quad \forall \ k = 1, 2, \dots, n_1 \\ & (v) \ \sum_{s \in S} (f(s))^T D^{2,l}(s) x(s) \leq \xi_l^2, \quad \forall \ l = 1, 2, \dots, n_2 \\ & (vi) \ \sum_{a^1 \in A^1(s)} f(s, a^1) = 1, \quad \forall \ s \in S \\ & (vii) \ f(s, a^1) \geq 0, \quad \forall \ s \in S, \ a^1 \in A^1(s) \end{aligned}$$

$$(viii) \ x(s, a^2) \geq 0, \ \forall \ s \in S, \ a^2 \in A^2(s)$$

$$(ix) \ \delta_k^1 \geq 0, \ \forall \ k = 1, 2, \dots, n_1$$

$$(x) \ \delta_l^2 \geq 0, \ \forall \ l = 1, 2, \dots, n_2.$$

with $\Phi(\eta^*) = 0$.

(b) If $\eta^* = (u^{*T}, z^{*T}, f^{*T}, x^{*T}, (\delta^{1*})^T, (\delta^{2*})^T)^T$ is a global minimum of [MP2] with $\Phi(\eta^*) = 0$, then, (f^*, g^*) is a Nash equilibrium of the constrained stochastic game G^c with discounted cost criterion, where

$$g^*(s, a^2) = \frac{x^*(s, a^2)}{\sum_{a^2 \in A^2(s)} x^*(s, a^2)}$$

for all $s \in S, a^2 \in A^2(s)$ whenever the denominator is non-zero (when it is zero $g^*(s)$ is chosen arbitrarily from $\wp(A^2(s))$).

Proof We can prove this by using the best response linear programs (21), (22), (23), (24) and with similar argument given in the proof of Theorem 1.

Remark 2 Similar to [MP1], [MP2] is also a non-convex constrained optimization problem.

Remark 3 Both [MP1] and [MP2] can be obtained from single mathematical program [MP4] given in Appendix (A).

2.2.3 Special cases

We consider two special cases. First, we consider nonzero sum game as defined in Section 2 with discounted cost criterion where the constraints of player 2 do not depend on the strategies of the player 1. Next, we briefly consider the zero sum game as considered in [1].

(i) Quadratic program in case of decoupled constraints

When the constraints of player 2 do not depend on the strategies of player 1, i.e., under condition (18) the mathematical program [MP2] reduces to a quadratic program [QP2] given below

$$\begin{aligned} [\text{QP2}] \quad \min_{\eta} \quad & \left[(f^T C^1 x - (\mathbf{1}_{|S|}^T z - (\delta^1)^T \xi^1)) \right. \\ & \left. + (f^T C^2 x - ((1 - \beta)\gamma^T u - (\delta^2)^T \xi^2)) \right] \end{aligned}$$

s.t.

$$\begin{aligned} (i) \ u(s) \leq & [(f(s))^T C^2(s)]_{a^2} + \sum_{l=1}^{n_2} \delta_l^2 d^{2,l}(s, a^2) \\ & + \beta \sum_{s' \in S} p(s'|s, a^2) u(s'), \ \forall \ s \in S, \ a^2 \in A^2(s) \end{aligned}$$

$$\begin{aligned}
(ii) \quad & z(s) \leq [C^1(s)x(s)]_{a^1} + \sum_{k=1}^{n_1} \delta_k^1 d_{sub}^{1,k}(s, a^1), \quad \forall s \in S, a^1 \in A^1(s) \\
(iii) \quad & \sum_{(s, a^2) \in \mathcal{K}^2} [\delta(s, s') - \beta p(s'|s, a^2)] x(s, a^2) = (1 - \beta)\gamma(s'), \quad \forall s' \in S \\
(iv) \quad & \sum_{(s, a^1) \in \mathcal{K}^1} d_{sub}^{1,k}(s, a^1) f(s, a^1) \leq \xi_k^1, \quad \forall k = 1, 2, \dots, n_1 \\
(v) \quad & \sum_{(s, a^2) \in \mathcal{K}^2} d^{2,l}(s, a^2) x(s, a^2) \leq \xi_l^2, \quad \forall l = 1, 2, \dots, n_2 \\
(vi) \quad & \sum_{a^1 \in A^1(s)} f(s, a^1) = 1, \quad \forall s \in S \\
(vii) \quad & f(s, a^1) \geq 0, \quad \forall s \in S, a^1 \in A^1(s) \\
(viii) \quad & x(s, a^2) \geq 0, \quad \forall s \in S, a^2 \in A^2(s) \\
(ix) \quad & \delta_k^1 \geq 0, \quad \forall k = 1, 2, \dots, n_1 \\
(x) \quad & \delta_l^2 \geq 0, \quad \forall l = 1, 2, \dots, n_2.
\end{aligned}$$

(ii) *Zero sum single controller constrained stochastic games*

The zero sum single controller constrained stochastic game with discounted cost criterion is considered in [1]. In [1], the first player has subscription based constraints and second player has realization based constraints which do not depend on the strategies of first player and these games can be solved by a linear program. Setting $C^1(s) = -C^2(s) = C(s)$ for all $s \in S$ the quadratic program [QP2] can be separated into primal-dual pair of linear programs which are same as given in [1].

2.3 A Numerical Example

We give one numerical example where immediate costs of player 2 corresponding to his constraints do not depend on the actions of player 1. We compute the Nash equilibrium of this game by solving corresponding quadratic program. The components of the stochastic game are

1. The state space $S = \{1, 2\}$.
2. The action sets of both the players are $A^i(s) = \{1, 2\}$, $i = 1, 2$, $s = 1, 2$.
3. The immediate costs of both the players that defines their expected cost which they want to minimize and transition probabilities of the game are given in the Table 1(a) and 1(b).

Table 1 Immediate costs and transition probabilities

(a) $s = 1$		(b) $s = 2$	
$(5,4)$	$(\frac{1}{2}, \frac{1}{2})$	$(2,3)$	$(1,0)$
$(7,3)$	$(\frac{1}{2}, \frac{1}{2})$	$(3,1)$	$(1,0)$
$(6,3)$	$(\frac{1}{3}, \frac{2}{3})$	$(4,2)$	$(\frac{1}{5}, \frac{4}{5})$
$(4,6)$	$(\frac{1}{3}, \frac{2}{3})$	$(3,4)$	$(\frac{1}{5}, \frac{4}{5})$

In both the tables above, the entry in upper triangle in each box gives the transition probabilities and the entry in lower triangle gives the immediate cost of both the players corresponding to the actions chosen by both the players in that state. For example, at state 1 when both the player choose action 1, then, player 1 gets immediate cost 5 and player 2 gets 4 and this is represented by entry $(5,4)$ and game will remain in state 1 with probability $\frac{1}{2}$ and it can move to state 2 with probability $\frac{1}{2}$ and this is represented by entry $(\frac{1}{2}, \frac{1}{2})$ in the table corresponding to state 1. It is easy to check that transition probabilities given in tables above satisfies the ergodicity assumption (A1).

4. Both the players have one constraint, i.e., player 1 has one subscription based constraint and player 2 has one realization based constraint. The subscription cost of player 1 and immediate cost of player 2 corresponding to each state-action pair are given in Table 2(a) and 2(b) respectively.

Table 2 Costs defining constraints

(a) $d_{sub}^1(s, a^1)$			(b) $d^2(s, a^2)$		
	$s = 1$	$s = 2$		$s = 1$	$s = 2$
$a^1 = 1$	2	3	$a^2 = 1$	1	4
$a^1 = 2$	3	1	$a^2 = 2$	2	5

5. The bound defining constraints are $\xi^1 = 4, \xi^2 = 2.5$.

From Table 1(a) and 1(b) it is clear that the game is controlled by player 2.

- (i) For average cost criterion we solve the quadratic program [QP1] corresponding to the above data, by using MATLAB and obtain

$$\eta^* = (3.0278, 4.1667, 2.833, 3.8667, 1.3067, 0.6944, 0.3056, 0.3472, 0.6528, 0.2667, 0.36, 0.3733, 0, 0.1867, 0).$$

Note that at η^* the objective function value is zero and hence it is the global minimum of quadratic program. We have $x^*(1, 1) = 0.2667$, $x^*(1, 2) = 0.36$, $x^*(2, 1) = 0.3733$, $x^*(2, 2) = 0$. From (10) we have $g^*(1, 1) = 0.4256$, $g^*(1, 2) = 0.5744$, $g^*(2, 1) = 1$, $g^*(2, 2) = 0$. From Theorem 1(b) the Nash equilibrium of constrained stochastic game defined above with average cost criterion is

$$f^* = ((0.6944, 0.3056), (0.3472, 0.6528)), \quad g^* = ((0.4256, 0.5744), (1, 0))$$

and the average costs of both the players at Nash equilibrium (f^*, g^*) are

$$\begin{aligned} C_{ea}^1(f^*, g^*) &= 4.4268 \\ C_{ea}^2(f^*, g^*) &= 3.0279. \end{aligned}$$

- (ii) For discounted cost criterion we take $\beta = 0.5$, $\gamma = (0.5, 0.5)$. We solve the quadratic program [QP2] corresponding to the above data, by using MATLAB and obtain

$$\eta^* = (10.2222, 10.8888, 3.5833, 1.4583, 1, 0, 0.5, 0.5, 0.3333, 0.25, 0.4167, 0, 0.2083, 0.9444).$$

Note that at η^* the objective function is zero and hence it is the global minimum of quadratic program. We have $x^*(1, 1) = 0.3333$, $x^*(1, 2) = 0.25$, $x^*(2, 1) = 0.4167$, $x^*(2, 2) = 0$. From (20) we have $g^*(1, 1) = 0.5714$, $g^*(1, 2) = 0.4286$, $g^*(2, 1) = 1$, $g^*(2, 2) = 0$. From Theorem 2(b) the Nash equilibrium of constrained stochastic game defined above with discounted cost criterion is

$$f^* = ((1, 0), (0.5, 0.5)), \quad g^* = ((0.5714, 0.4286), (1, 0))$$

and the discounted costs of both the players at Nash equilibrium (f^*, g^*) are

$$\begin{aligned} C_\beta^1(\gamma, f^*, g^*) &= 4.2082 \\ C_\beta^2(\gamma, f^*, g^*) &= 2.9166. \end{aligned}$$

3 Constrained stochastic game with independent state processes

In this section we consider a N -player constrained stochastic game with independent state processes as discussed in [2]. In these games each player controls his own Markov chain, whose transition probabilities do not depend on the states and actions of other players. In these games at any time, each player has information only about current and past states of his Markov chain as well as of his previous actions and does not have any information about the states and actions of other players. However, each player wants to minimize his expected average cost that depends on the strategies of all the players. The expected average constraints of each player also depend on the strategies of all the players. These games come under the class of decentralized stochastic games.

The game is described by the tuple $(S^i, \gamma^i, A^i, c^i, d^i, p^i, \xi^i)$, $i = 1, 2, \dots, N$, where:

- (i) S^i is the finite state space of player i , $i = 1, \dots, N$. The generic element of S^i is denoted by s^i . Define, $S := \times_{j=1}^N S^j$ and $S^{-i} := \times_{j \neq i} S^j$ (\times stands for the product space). The element of S is denoted by s where $s = (s^1, s^2, \dots, s^N)$ and $s^{-i} \in S^{-i}$ denote the vector of states s^j , $j \neq i$.

- (ii) γ^i is the probability distribution for the initial state of player i , $i = 1, \dots, N$. We assume that the initial states of all the players are independent. Denote $\gamma = (\gamma^1, \gamma^2, \dots, \gamma^N)$.
- (iii) A^i is the finite action (strategy) set of player i and its element is denoted by a^i , $i = 1, \dots, N$. $A^i(s^i)$ denotes the set of all actions of player i at state s^i and $A^i = \bigcup_{s^i \in S^i} A^i(s^i)$. We denote $a = (a^1, a^2, \dots, a^N)$ and a^{-i} as vector of actions a^j , $j \neq i$.
- (iv) Define, $\mathcal{K}^i = \{(s^i, a^i) | s^i \in S^i, a^i \in A^i(s^i)\}$, $i = 1, 2, \dots, N$ and $\mathcal{K} = \times_{i=1}^N \mathcal{K}^i$, $\mathcal{K}^{-i} = \times_{j \neq i} \mathcal{K}^j$.
- (v) $c^i : \mathcal{K} \rightarrow \mathbb{R}$ is immediate cost of player i , $i = 1, \dots, N$. Specifically, $c^i(s, a)$ is the immediate cost incurred by player i , $i = 1, 2, \dots, N$, when state of players is (s^1, s^2, \dots, s^N) and actions chosen by them are (a^1, a^2, \dots, a^N) respectively. Each player i , $i = 1, 2, \dots, N$, wants to minimize the expected average cost involving $c^i(\cdot, \cdot)$.
- (vi) $d^i = (d^{i,1}, d^{i,2}, \dots, d^{i,n_i})$, where $d^{i,k} : \mathcal{K} \rightarrow \mathbb{R}$ for all $k = 1, 2, \dots, n_i$ are immediate costs of player i , $i = 1, \dots, N$. These $d^{i,k}(\cdot, \cdot)$ are involved in the k th constraint, $k = 1, 2, \dots, n_i$, on expected average cost of player i , $i = 1, \dots, N$.
- (vii) $p^i : \mathcal{K}^i \rightarrow \wp(S^i)$ is the transition probability of player i , $i = 1, \dots, N$, where $p^i(\bar{s}^i | s^i, a^i)$ is the probability that the state of player i moves from state s^i to \bar{s}^i if he chooses action $a^i \in A^i(s^i)$.
- (viii) $\xi^i = (\xi_1^i, \xi_2^i, \dots, \xi_{n_i}^i)$ are the bounds defining the constraints of player i , $i = 1, \dots, N$.

The game dynamics are as follows. Initially, at time $t = 0$ state of the game is $s = (s^1, s^2, \dots, s^N)$ where $s^i \in S^i$ is chosen according to independent random variables γ^i , $i = 1, 2, \dots, N$. Players independently choose actions $a = (a^1, a^2, \dots, a^N)$ with $a^i \in A^i(s^i)$, $i = 1, 2, \dots, N$. Player i obtains an immediate cost $c^i(s, a)$, $i = 1, 2, \dots, N$. Apart from this cost, player i , $i = 1, 2, \dots, N$, also receives another n_i costs $\{d^{i,k}(s, a)\}$, $k = 1, 2, \dots, n_i$. These $\{d^{i,k}(\cdot, \cdot)\}$, $k = 1, 2, \dots, n_i$, are involved in the expected average cost functionals of player i which are constrained by specified bounds $\{\xi_k^i\}$, $k = 1, \dots, n_i$. Now, the state of player i switches to a new state \bar{s}^i at time $t = 1$ with probability $p^i(\bar{s}^i | s^i, a^i)$, $i = 1, \dots, N$. At time $t = 1$, in state \bar{s}^i , player i then independently chooses an action \bar{a}^i , receives costs $c^i(\bar{s}, \bar{a})$ and $\{d^{i,k}(\bar{s}, \bar{a})\}$, $k = 1, \dots, n_i$ and $i = 1, \dots, N$. The next state for this player is \tilde{s}^i with probability $p^i(\tilde{s}^i | \bar{s}^i, \bar{a}^i)$. The dynamics of the Markov chains repeat at new state $\tilde{s} = (\tilde{s}^1, \dots, \tilde{s}^N)$ and game continues for infinite time horizon.

While transition probabilities depend only on the present state and action used, actions that are used can depend on ‘past’, as in history dependent strategies. Define a history of player i , $i = 1, 2, \dots, N$, at time t as $h_t^i = (s_0^i, a_0^i, s_1^i, a_1^i, \dots, s_{t-1}^i, a_{t-1}^i, s_t^i)$ where $s_t^i \in S^i$, $a_t^i \in A^i(s_t^i)$, $i = 1, 2, \dots, N$, $t = 0, 1, 2, \dots$. Let H_t^i denote the set of all possible histories of length t of player i . Each player observes his own history and does not have any information about the other player’s history. A decision rule $f_t^i : H_t^i \rightarrow \wp(A^i(s_t^i))$ of player i at time t is a function which assigns to each history of length t of player i , a

probability measure over action set of player i . This means that under decision rule f_t^i player i chooses action a^i with probability $f_t^i(h_t^i, a^i)$. The sequence of decision rules is called the strategy of the player. Let $f^{ih} = (f_0^i, f_1^i, \dots, f_t^i, \dots)$ denote the strategy of player i , $i = 1, 2, \dots, N$, and is called history dependent (behavioral) strategy. Note that the strategies of players do not depend on the realizations of the costs. If strategies were allowed to depend on such costs, then a player could use the costs to estimate the state and actions of the other players.

Let F^i denote the set of all history dependent strategies of player i and $F = \times_{i=1}^N F^i$ be the class of history dependent multi-strategies. These strategies are called Markovian if at every decision epoch the decision rule depends only on the current state but the decision rule can differ at every epoch. A stationary strategy is a Markovian strategy which is independent of the time, i.e., at every decision epoch the decision rule is same. So, for stationary strategy $f_t^i = f^i$ for all t , i.e., (f^i, f^i, f^i, \dots) is a stationary strategy of player i . We denote, with some abuse of notations, f^i as the stationary strategy of player i . Let F_{S^i} denote the set of all stationary strategies of player i and $F_S = \times_{i=1}^N F_{S^i}$ denote the class of stationary multi-strategies. For, $i = 1, 2, \dots, N$, stationary strategy $f^i \in F_{S^i}$ is identified with $f^i = ((f^i(1))^T, (f^i(2))^T, \dots, (f^i(|S^i|))^T)^T$, where $f^i(s^i) = (f^i(s^i, 1), f^i(s^i, 2), \dots, f^i(s^i, |A^i(s^i)|))^T$ for all $s^i \in S^i$. For all $s^i \in S^i$, $f^i(s^i, a^i)$ is then, the probability of choosing action $a^i \in A^i(s^i)$ by player i , $i = 1, \dots, N$. For $f^h \in F$ we denote f^{-ih} as the vector of strategies f^{jh} , $j \neq i$, and for any $g^{ih} \in F^i$ we define (f^{-ih}, g^{ih}) to be the multi-strategy, where, for $j \neq i$, player j uses f^{jh} and player i uses g^{ih} . Under mild assumptions, which we also make, Altman, *et al* [2] show that a Nash equilibrium exists for the above constrained stochastic game within the class of stationary strategies.

This leads to the introduction of vector stochastic process $\{X_t, \mathbb{A}_t\}_{t=0}^\infty$, where $X_t = (X_t^1, X_t^2, \dots, X_t^N)$, $\mathbb{A}_t = (\mathbb{A}_t^1, \mathbb{A}_t^2, \dots, \mathbb{A}_t^N)$, X_t^i denote the state of the player i and \mathbb{A}_t^i denote the action chosen by player i at time t , $t = 0, 1, \dots$. An initial distribution γ together with multi-strategy $f^h \in F$ defines a unique probability measure $\mathbb{P}_{f^h}^\gamma$ on an appropriate probability space with respect to which the laws of vector stochastic process $\{X_t, \mathbb{A}_t\}_{t=0}^\infty$ of states and actions can be defined. The expectation operator on this probability space is denoted by $\mathbb{E}_{f^h}^\gamma$.

The expected average costs

These costs are average functionals of states and actions of all the players and each player minimizes his cost functionals. For given initial distribution γ and multi-strategy f^h the expected average cost of player i , $i = 1, 2, \dots, N$ is defined as

$$C_{ea}^i(\gamma, f^h) = \limsup_{T \rightarrow \infty} \frac{1}{T} \sum_{t=0}^{T-1} \mathbb{E}_{f^h}^\gamma c^i(X_t, \mathbb{A}_t). \quad (25)$$

The expected average constraints

The constraints of each player are defined by average functionals of states and actions of all the players which are bounded by given reals. For given initial distribution γ and multi-strategy f^h the expected average costs of player i , $i = 1, 2, \dots, N$ are defined as

$$D_{ea}^{i,k}(\gamma, f^h) = \limsup_{T \rightarrow \infty} \frac{1}{T} \sum_{t=0}^{T-1} \mathbb{E}_{f^h}^\gamma d^{i,k}(X_t, \mathbb{A}_t), \quad \forall k = 1, 2, \dots, n_i.$$

$D_{ea}^{i,k}(\cdot, \cdot)$ can capture the average consumption of resource k , $k = 1, 2, \dots, n_i$, by player i , $i = 1, 2, \dots, N$. The constraints of player i , $i = 1, 2, \dots, N$ are given as

$$D_{ea}^{i,k}(\gamma, f^h) \leq \xi_k^i, \quad \forall k = 1, 2, \dots, n_i. \quad (26)$$

The constraints (26) captures the fact that average consumption of resource k by player i , $i = 1, 2, \dots, N$, when player i uses strategy f^{ih} and other players use f^{-ih} is not more than given ξ_k^i , $k = 1, 2, \dots, n_i$.

All the players choose their strategies independently and want to minimize their expected average cost from (25) subject to their constraints from (26). We denote this constrained stochastic game by G_{ea}^c . The multi-strategy $f^h = (f^{1h}, f^{2h}, \dots, f^{Nh})$ is called i -feasible if it satisfies i th player's constraints from (26) and it is called feasible if it is i -feasible for every $i = 1, 2, \dots, N$. Let F^ξ denote the set of all feasible history dependent multi-strategies and F_S^ξ denote the set of all stationary feasible multi-strategies for the constrained stochastic game G_{ea}^c . We shall assume throughout that F_S^ξ is non-empty.

We now recall the definition of Nash equilibrium as given in [2]. A multi-strategy $f^{h*} \in F^\xi$ is called the Nash equilibrium of the constrained stochastic game G_{ea}^c , if for each player $i = 1, 2, \dots, N$ and for any f^{ih} such that (f^{ih}, f^{-ih*}) is i -feasible, one has that

$$C_{ea}^i(\gamma, f^{h*}) \leq C_{ea}^i(\gamma, f^{ih}, f^{-ih*}).$$

Thus, unilateral deviation by any player i , $i = 1, \dots, N$ from equilibrium strategy f^{h*} is not possible, because in that case, either at least one of his constraints will be violated or it will result in a cost for player i that is not lower than the one achieved by feasible equilibrium strategy f^{h*} . A stationary multi-strategy $f^* \in F_S^\xi$ is said to be Nash equilibrium of constrained stochastic game G_{ea}^c , if for each player $i = 1, 2, \dots, N$ and for any f^i such that (f^i, f^{-i*}) is i -feasible, one has that

$$C_{ea}^i(\gamma, f^*) \leq C_{ea}^i(\gamma, f^i, f^{-i*}).$$

This can be seen by noticing that when all players j , $j \neq i$, fix their strategy as a stationary strategy then player i is faced with a constrained Markov decision process (CMDP) where optimal strategy always exists in the space of stationary strategies [20].

Assumptions [Altman, et al. [2]]

As similar to [2] we also have the following assumptions:

- (A1) Ergodicity: For each player i , $i = 1, \dots, N$, and for any stationary strategy f^i the state process of player i is an irreducible Markov chain with one ergodic class (and possibly some transient states).
- (A2) Strong Slater condition: Every player i , $i = 1, \dots, N$ has some strategy g^i such that for any multi-strategy f^{-i} of other players,

$$D_{ea}^{i,k}(\gamma, (f^{-i}, g^i)) < \xi_k^i, \quad \forall k = 1, 2, \dots, n_i.$$

- (A3) The players do not observe their costs, i.e., the strategy chosen by any player does not depend on the realization of the cost.

The last assumption is due to the definition of the strategies. If strategies were allowed to depend on the realization of the costs, then a player can use the cost to estimate other player's states and actions. As the Nash equilibrium exists in stationary strategies under the assumptions (A1)-(A3) [2], from now onwards we restrict ourselves to the class of stationary strategies.

3.1 Average occupation measure

For each player i , $i = 1, 2, \dots, N$, using a stationary strategy f^i and initial distribution γ^i define the average occupation measure as

$$\pi_{ea}^i(\gamma^i, f^i) := \{ \pi_{ea}^i(\gamma^i, f^i; s^i, a^i) : s^i \in S^i, a^i \in A^i(s^i) \}.$$

For all $s^i \in S^i$, $a^i \in A^i(s^i)$, $\pi_{ea}^i(\gamma^i, f^i; s^i, a^i)$ is given by

$$\pi_{ea}^i(\gamma^i, f^i; s^i, a^i) = \pi^{f^i}(s^i) f^i(s^i, a^i) \quad (27)$$

where $\pi^{f^i} = (\pi^{f^i}(1), \pi^{f^i}(2), \dots, \pi^{f^i}(|S^i|))$ is the unique steady state distribution of Markov chain induced by strategy f^i of player i , which exists under (A1). $\pi_{ea}^i(\gamma^i, f^i)$ can be considered as probability measure over \mathcal{K}^i that assigns probability $\pi_{ea}^i(\gamma^i, f^i; s^i, a^i)$ to state-action pair (s^i, a^i) . The occupation measure defined in (27) is unique and independent from initial distribution γ^i , so, we drop γ^i from the notation. For any multi-strategy $f \in F_S$ the expected average costs for each player i , $i = 1, 2, \dots, N$, can be written in terms of occupation measure as

$$C_{ea}^i(f) = \sum_{(s,a) \in \mathcal{K}} \left[\prod_{j=1}^N \pi_{ea}^j(f^j; s^j, a^j) \right] c^i(s, a).$$

$$D_{ea}^{i,k}(f) = \sum_{(s,a) \in \mathcal{K}} \left[\prod_{j=1}^N \pi_{ea}^j(f^j; s^j, a^j) \right] d^{i,k}(s, a), \quad \forall k = 1, 2, \dots, n_i.$$

Let $Q_{ea}^i, i = 1, 2, \dots, N$, be the set of vectors $x^i \in \mathbb{R}^{|\mathcal{K}^i|}$ satisfying

$$\begin{cases} \sum_{(s^i, a^i) \in \mathcal{K}^i} (\delta(s^i, \bar{s}^i) - p^i(\bar{s}^i | s^i, a^i)) x^i(s^i, a^i) = 0, \quad \forall \bar{s}^i \in S^i \\ \sum_{(s^i, a^i) \in \mathcal{K}^i} x^i(s^i, a^i) = 1 \\ x^i(s^i, a^i) \geq 0, \quad \forall s^i \in S^i, a^i \in A^i(s^i). \end{cases}$$

The space of stationary strategies is complete, i.e., the set of occupation measures achieved by history dependent strategies equals to those achieved by stationary strategies and further equals to the set $Q_{ea}^i, i = 1, 2, \dots, N$ [20]. It is known that for each $(s^i, a^i) \in \mathcal{K}^i$, $x^i(s^i, a^i) = \pi_{ea}^i(f^i; s^i, a^i)$ where

$$f^i(s^i, a^i) = \frac{x^i(s^i, a^i)}{\sum_{a^i \in A^i(s^i)} x^i(s^i, a^i)} \quad (28)$$

whenever denominator is nonzero (when it is zero $f^i(s^i)$ is chosen arbitrarily from $\wp(A^i(s^i))$) [20].

We use the following notations throughout this section. For $i = 1, 2, \dots, N$,

- $u^i = (u^i(1), u^i(2), \dots, u^i(|S^i|))^T$.
- $v^i \in \mathbb{R}$.
- $x^i = ((x^i(1))^T, (x^i(2))^T, \dots, (x^i(|S^i|))^T)^T$.
- $x^i(s^i) = (x^i(s^i, 1), x^i(s^i, 2), \dots, x^i(s^i, |A^i(s^i)|))^T$.
- $\delta^i = (\delta_1^i, \delta_2^i, \dots, \delta_{n_i}^i)^T$.

The costs of player $i, i = 1, 2, \dots, N$, when he uses action a^i at state s^i and other players use f^{-i} is defined as in [2],

$$\begin{aligned} c^i(f^{-i}; s^i, a^i) &= \sum_{(s, a)^{-i} \in \mathcal{K}^{-i}} \left[\prod_{j=1; j \neq i}^N \pi_{ea}^j(f^j; s^j, a^j) \right] c^i(s, a). \\ d^{i,k}(f^{-i}; s^i, a^i) &= \sum_{(s, a)^{-i} \in \mathcal{K}^{-i}} \left[\prod_{j=1; j \neq i}^N \pi_{ea}^j(f^j; s^j, a^j) \right] d^{i,k}(s, a), \quad \forall k = 1, 2, \dots, n_i. \end{aligned}$$

3.2 Mathematical programming formulation

We show the one to one correspondence between the stationary Nash equilibria of this game and the global minima of one mathematical program.

Best response linear programs

The best response of each player $i, i = 1, 2, \dots, N$, against fixed stationary strategy f^{-i} of other players is given by solving a constrained Markov decision model, which, in turn, can be obtained by a linear program in our setting [20].

The best response of player i against fixed strategy f^{-i} of other players is given by the linear program below:

$$\left. \begin{array}{l} \min_{x^i} \sum_{(s^i, a^i) \in \mathcal{K}^i} c^i(f^{-i}; s^i, a^i) x^i(s^i, a^i) \\ \text{s.t.} \\ (i) \sum_{(s^i, a^i) \in \mathcal{K}^i} (\delta(s^i, \bar{s}^i) - p^i(\bar{s}^i | s^i, a^i)) x^i(s^i, a^i) = 0, \quad \forall \bar{s}^i \in S^i \\ (ii) \sum_{(s^i, a^i) \in \mathcal{K}^i} x^i(s^i, a^i) = 1 \\ (iii) \sum_{(s^i, a^i) \in \mathcal{K}^i} d^{i,k}(f^{-i}; s^i, a^i) x^i(s^i, a^i) \leq \xi_k^i, \quad \forall k = 1, 2, \dots, n_i \\ (iv) x^i(s^i, a^i) \geq 0, \quad \forall s^i \in S^i, a^i \in A^i(s^i) \end{array} \right\} \quad (29)$$

If x^{i*} is the optimal solution of the linear program (29), then, the best response f^{i*} of player i can be obtained from (28) [20]. The dual of linear program (29) is

$$\left. \begin{array}{l} \max_{v^i, u^i, \delta^i} \left[v^i - \sum_{k=1}^{n_i} \delta_k^i \xi_k^i \right] \\ \text{s.t.} \\ (i) v^i + u^i(s^i) \leq c^i(f^{-i}; s^i, a^i) + \sum_{k=1}^{n_i} d^{i,k}(f^{-i}; s^i, a^i) \delta_k^i \\ \quad + \sum_{\bar{s}^i \in S^i} p^i(\bar{s}^i | s^i, a^i) u^i(\bar{s}^i), \quad \forall s^i \in S^i, a^i \in A^i(s^i) \\ (ii) \delta_k^i \geq 0, \quad \forall k = 1, 2, \dots, n_i. \end{array} \right\} \quad (30)$$

By using N primal-dual pair of linear programs given by (29), (30), we show the one to one correspondence between the stationary Nash equilibria of constrained stochastic game G_{ea}^c and global minima of a mathematical program [MP3]. Let $\zeta^T := (v^i, (u^i)^T, (x^i)^T, (\delta^i)^T)_{i=1}^N$ and $\psi(\zeta)$ denote the decision variables and the objective function of [MP3] respectively. ζ^T is a $1 \times (N + \sum_{i=1}^N |S^i| + \sum_{i=1}^N \sum_{s^i \in S^i} |A^i(s^i)| + \sum_{i=1}^N n_i)$ dimensional vector.

Theorem 3 (a) *If $(f^{i*})_{i=1}^N$ is a Nash equilibrium of the constrained stochastic game G_{ea}^c , then, there exists a vector $\zeta^{*T} = (v^{i*}, (u^{i*})^T, (x^{i*})^T, (\delta^{i*})^T)_{i=1}^N$ such that it is a global minimum of mathematical program [MP3] given below*

$$\begin{aligned} \text{[MP3]} \quad & \min_{\zeta} \sum_{i=1}^N \left[\sum_{(s,a) \in \mathcal{K}} \left(\prod_{j=1}^N x^j(s^j, a^j) \right) c^i(s, a) - \left(v^i - \sum_{k=1}^{n_i} \delta_k^i \xi_k^i \right) \right] \\ \text{s.t.} \quad & \end{aligned}$$

$$\begin{aligned}
(i) \quad & v^i + u^i(s^i) \leq \sum_{(s,a)^{-i} \in \mathcal{K}^{-i}} \left(\prod_{j=1; j \neq i}^N x^j(s^j, a^j) \right) c^i(s, a) \\
& + \sum_{k=1}^{n_i} \delta_k^i \left[\sum_{(s,a)^{-i} \in \mathcal{K}^{-i}} \left(\prod_{j=1; j \neq i}^N x^j(s^j, a^j) \right) d^{i,k}(s, a) \right] \\
& + \sum_{\bar{s}^i \in S^i} p^i(\bar{s}^i | s^i, a^i) u^i(\bar{s}^i), \quad \forall s^i \in S^i, a^i \in A^i(s^i), i = 1, 2, \dots, N \\
(ii) \quad & \sum_{(s^i, a^i) \in \mathcal{K}^i} (\delta(s^i, \bar{s}^i) - p^i(\bar{s}^i | s^i, a^i)) x^i(s^i, a^i) = 0, \quad \forall \bar{s}^i \in S^i, i = 1, 2, \dots, N \\
(iii) \quad & \sum_{(s^i, a^i) \in \mathcal{K}^i} x^i(s^i, a^i) = 1, \quad \forall i = 1, 2, \dots, N \\
(iv) \quad & \sum_{(s,a) \in \mathcal{K}} \left(\prod_{j=1}^N x^j(s^j, a^j) \right) d^{i,k}(s, a) \leq \xi_k^i, \quad \forall k = 1, 2, \dots, n_i, i = 1, 2, \dots, N \\
(v) \quad & x^i(s^i, a^i) \geq 0, \quad \forall s^i \in S^i, a^i \in A^i(s^i), i = 1, 2, \dots, N \\
(vi) \quad & \delta_k^i \geq 0, \quad \forall k = 1, 2, \dots, n_i, i = 1, 2, \dots, N.
\end{aligned}$$

with $\psi(\zeta^*) = 0$.

- (b) If $\zeta^{*T} = (v^{i*}, (u^{i*})^T, (x^{i*})^T, (\delta^{i*})^T)_{i=1}^N$ is a global minimum of [MP3] with $\psi(\zeta^*) = 0$ then $(f^{i*})_{i=1}^N$ is a Nash equilibrium of the constrained stochastic game G_{ea}^c where,

$$f^{i*}(s^i, a^i) = \frac{x^{i*}(s^i, a^i)}{\sum_{a^i \in A^i(s^i)} x^{i*}(s^i, a^i)}$$

for all $s^i \in S^i, a^i \in A^i(s^i), i = 1, 2, \dots, N$ whenever the denominator is non-zero (when it is zero $f^{i*}(s^i)$ is chosen arbitrarily from $\wp(A^i(s^i))$).

Proof (a) Let $(f^{i*})_{i=1}^N$ be a Nash equilibrium of the constrained stochastic game G_{ea}^c . For each $i = 1, 2, \dots, N$, we construct occupation measures x^{i*} as in (27) corresponding to stationary strategies f^{i*} . Then, the constraints in (ii), (iii) and (v) are satisfied by $(x^{i*})_{i=1}^N$. The multi-strategy $(f^{i*})_{i=1}^N$ is feasible because it is a Nash equilibrium, so the constraints in (iv) are also satisfied by $(x^{i*})_{i=1}^N$. For each $i = 1, 2, \dots, N$, f^{i*} is best response of player i against fixed strategy f^{-i*} of other players; so, x^{i*} as constructed above will be optimal solution of linear program (29) for this fixed f^{-i*} from Proposition 3.1(ii) of [2]. From strong duality theorem [24], [25] there exist optimal solution $(v^{i*}, u^{i*}, \delta^{i*})$ of (30) such that the constraints in (i) and (vi) of [MP3] are satisfied by $(v^{i*}, u^{i*}, x^{i*}, \delta^{i*})_{i=1}^N$ and objective function value of (29) and (30) are same. In other words we have a point $\zeta^{*T} = (v^{i*}, (u^{i*})^T, (x^{i*})^T, (\delta^{i*})^T)_{i=1}^N$ which is feasible for [MP3] and

$$\sum_{(s,a) \in \mathcal{K}} \left(\prod_{j=1}^N x^{j*}(s^j, a^j) \right) c^i(s, a) = v^{i*} - \sum_{k=1}^{n_i} \delta_k^{i*} \xi_k^i, \quad \forall i = 1, 2, \dots, N.$$

From the construction of the objective function, $\psi(\zeta^*) = 0$.

Let ζ be any feasible point of [MP3]. For each $i = 1, 2, \dots, N$, multiply each constraint in (i) of [MP3] corresponding to pair $(s^i, a^i) \in \mathcal{K}^i$ by $x^i(s^i, a^i)$, add over all $(s^i, a^i) \in \mathcal{K}^i$ and by then using the constraints (ii)-(vi) we have

$$\sum_{(s,a) \in \mathcal{K}} \left(\prod_{j=1}^N x^j(s^j, a^j) \right) c^i(s, a) \geq v^i - \sum_{k=1}^{n_i} \delta_k^i \xi_k^i, \quad \forall i = 1, 2, \dots, N. \quad (31)$$

From (31) we have $\psi(\zeta) \geq 0$ for all feasible points ζ of [MP3]. Thus ζ^* is a global minimum of the [MP3].

(b) Let ζ^* be a global minimum of [MP3] such that $\psi(\zeta^*) = 0$. As ζ^* is a feasible point of [MP3] then (31) will also hold for ζ^* , i.e.,

$$\sum_{(s,a) \in \mathcal{K}} \left(\prod_{j=1}^N x^{j*}(s^j, a^j) \right) c^i(s, a) \geq v^{i*} - \sum_{k=1}^{n_i} \delta_k^{i*} \xi_k^i, \quad \forall i = 1, 2, \dots, N.$$

From above we see that all N terms of the objective function are non-negative at ζ^* . But at ζ^* the objective function value is zero which means that all the terms are individually zero, i.e.,

$$\sum_{(s,a) \in \mathcal{K}} \left(\prod_{j=1}^N x^{j*}(s^j, a^j) \right) c^i(s, a) = v^{i*} - \sum_{k=1}^{n_i} \delta_k^{i*} \xi_k^i, \quad \forall i = 1, 2, \dots, N. \quad (32)$$

Fix ζ^* and for each $i = 1, 2, \dots, N$, multiply each constraint in (i) corresponding to pair $(s^i, a^i) \in \mathcal{K}^i$ by $x^i(s^i, a^i)$ and add over all $(s^i, a^i) \in \mathcal{K}^i$ and by using the constraints (ii)-(vi) and (32) we have for each $i = 1, 2, \dots, N$

$$\sum_{(s,a) \in \mathcal{K}} \left(\prod_{j=1}^N x^{j*}(s^j, a^j) \right) c^i(s, a) \leq \sum_{(s,a) \in \mathcal{K}} x^i(s^i, a^i) \left(\prod_{j=1; j \neq i}^N x^{j*}(s^j, a^j) \right) c^i(s, a)$$

for all i -feasible (x^i, x^{-i*}) . In other words we can say that for each $i = 1, 2, \dots, N$

$$C_{ea}^i(f^*) \leq C_{ea}^i(f^i, f^{-i*}), \quad \forall i\text{-feasible } (f^i, f^{-i*}).$$

That is $(f^{i*})_{i=1}^N$ is Nash equilibrium of the constrained stochastic game G_{ea}^c where

$$f^{i*}(s^i, a^i) = \frac{x^{i*}(s^i, a^i)}{\sum_{a^i \in A^i(s^i)} x^{i*}(s^i, a^i)}$$

for all $s^i \in S^i$, $a^i \in A^i(s^i)$, $i = 1, 2, \dots, N$ whenever the denominator is non-zero (when it is zero $f^{i*}(s^i)$ is chosen arbitrarily from $\wp(A^i(s^i))$).

Remark 4 It is easy to see that [MP3] is also a non-convex constrained optimization problem.

3.2.1 Special cases

We consider two special cases. First, we consider two player nonzero sum constrained stochastic game as defined in Section 3, where, the constraints of both the players are decoupled. Next, we consider two player zero sum game as considered in [7].

(i) The case of two player game with decoupled constraints

Here we consider the situation where there are only two players and the constraints of each player do not depend on the strategies of the other player. This is possible when immediate costs of each player which correspond to his constraints do not depend on the state and actions of the other player, i.e., $d^{i,k}(s^1, s^2, a^1, a^2) = d^{i,k}(s^i, a^i)$ for all $s^i \in S^i$, $a^i \in A^i(s^i)$, $k = 1, 2, \dots, n_i$, $i = 1, 2$. We see that the mathematical program [MP3] reduces to a quadratic program [QP3] as given below

$$[\text{QP3}] \quad \min \sum_{i=1}^2 \left[\sum_{(s^1, a^1, s^2, a^2)} \left(\prod_{j=1}^2 x^j(s^j, a^j) \right) c^i(s^1, s^2, a^1, a^2) - \left(v^i - \sum_{k=1}^{n_i} \delta_k^i \xi_k^i \right) \right]$$

s.t.

$$\begin{aligned} (i) \quad & v^1 + u^1(s^1) \leq \sum_{(s^2, a^2) \in \mathcal{K}^2} c^1(s^1, s^2, a^1, a^2) x^2(s^2, a^2) + \sum_{k=1}^{n_1} d^{1,k}(s^1, a^1) \delta_k^1 \\ & + \sum_{\bar{s}^1 \in S^1} p^1(\bar{s}^1 | s^1, a^1) u^1(\bar{s}^1), \quad \forall s^1 \in S^1, a^1 \in A^1(s^1) \\ (ii) \quad & v^2 + u^2(s^2) \leq \sum_{(s^1, a^1) \in \mathcal{K}^1} c^2(s^1, s^2, a^1, a^2) x^1(s^1, a^1) + \sum_{k=1}^{n_2} d^{2,k}(s^2, a^2) \delta_k^2 \\ & + \sum_{\bar{s}^2 \in S^2} p^2(\bar{s}^2 | s^2, a^2) u^2(\bar{s}^2), \quad \forall s^2 \in S^2, a^2 \in A^2(s^2) \\ (iii) \quad & \sum_{(s^i, a^i) \in \mathcal{K}^i} (\delta(s^i, \bar{s}^i) - p^i(\bar{s}^i | s^i, a^i)) x^i(s^i, a^i) = 0, \quad \forall \bar{s}^i \in S^i, i = 1, 2 \\ (iv) \quad & \sum_{(s^i, a^i) \in \mathcal{K}^i} x^i(s^i, a^i) = 1, \quad \forall i = 1, 2 \\ (v) \quad & \sum_{(s^i, a^i) \in \mathcal{K}^i} d^{i,k}(s^i, a^i) x^i(s^i, a^i) \leq \xi_k^i, \quad \forall k = 1, 2, \dots, n_i, i = 1, 2 \\ (ix) \quad & x^i(s^i, a^i) \geq 0, \quad \forall s^i \in S^i, a^i \in A^i(s^i), i = 1, 2 \\ (x) \quad & \delta_k^i \geq 0, \quad \forall k = 1, 2, \dots, n_i, i = 1, 2. \end{aligned}$$

(ii) Zero sum constrained stochastic game [7]

As a further special case of constrained stochastic game G_{ea}^c we consider two player zero sum game with decoupled constraints [7]. This class of games with both unichain and multichain structure on the state processes of both the players can be solved by linear programs [7]. For zero sum case simply set $c^1(s^1, s^2, a^1, a^2) = -c^2(s^1, s^2, a^1, a^2) = c(s^1, s^2, a^1, a^2)$ for all $s^1 \in S^1, s^2 \in S^2, a^1 \in A^1(s^1), a^2 \in A^2(s^2)$ then the quadratic program [QP3] can be separated into a primal-dual pair of linear programs which are same as given in [7] in unichain case.

3.3 A numerical example

In this section we give one numerical example of a two player game where constraints of both the players are decoupled. We compute the Nash equilibrium of this game by solving quadratic program [QP3]. The components of the stochastic game are:

1. The state space of player 1 and player 2 are $S^1 = \{1, 2\}, S^2 = \{3, 4\}$ respectively.
2. The action sets of player 1 are $A^1(s^1) = \{1, 2\}$ for all $s^1 \in S^1$ and action sets of player 2 are $A^2(s^2) = \{1, 2\}$ for all $s^2 \in S^2$.
3. The immediate costs of both the players, which are involved in their expected average costs they want to minimize, are given in Tables 3(a), 3(b), 3(c) and 3(d). These tables summarize the immediate costs of both the

Table 3 Immediate costs

(a) $(s^1, s^2) = (1, 3)$				(b) $(s^1, s^2) = (1, 4)$			
$a^1 \backslash a^2$		$a^2 = 1$	$a^2 = 2$	$a^1 \backslash a^2$		$a^2 = 1$	$a^2 = 2$
$a^1 = 1$		(2,3)	(3,1)	$a^1 = 1$		(5,2)	(3,4)
$a^1 = 2$		(4,2)	(2,4)	$a^1 = 2$		(3,2)	(4,1)
(c) $(s^1, s^2) = (2, 3)$				(d) $(s^1, s^2) = (2, 4)$			
$a^1 \backslash a^2$		$a^2 = 1$	$a^2 = 2$	$a^1 \backslash a^2$		$a^2 = 1$	$a^2 = 2$
$a^1 = 1$		(3,5)	(4,6)	$a^1 = 1$		(4,5)	(3,1)
$a^1 = 2$		(5,2)	(2,1)	$a^1 = 2$		(1,2)	(4,3)

players in all the possible states. For example in Table 3(a) the entry (2, 3) represent 2 as immediate cost of player 1 when first player is in state 1 and he chooses action 1 and second player is in state 3 and chooses action 1. Similar explanation is for 3 and other entries in all the tables.

4. The transition probabilities of first and second Markov chains (one for each player) are given in the Tables 4(a) and 4(b) respectively. We can easily

check that both the Markov chains are unichain. In first Markov chain state 1 is transient for some strategies of player 1 and state 2 is recurrent for every strategy f^1 of player 1. In the second Markov chain both the states 3 and 4 are recurrent for every strategy f^2 of player 2. So, the assumption (A1) is satisfied.

Table 4 Transition probabilities of both the Markov chains

(a) $p^1(\cdot s^1, a^1)$			(b) $p^2(\cdot s^2, a^2)$		
	$a^1 = 1$	$a^1 = 2$		$a^2 = 1$	$a^2 = 2$
$s^1 = 1$	(0.5, 0.5)	(0.33, 0.67)	$s^2 = 3$	(0.67, 0.33)	(0.4, 0.6)
$s^1 = 2$	(1, 0)	(0, 1)	$s^2 = 4$	(0.25, 0.75)	(1, 0)

5. Each player has one constraint. The immediate costs of both the players which are used in their expected average constraints are given in Table 5(a) and 5(b).

Table 5 Immediate costs defining constraints

(a) $d^1(s^1, a^1)$			(b) $d^2(s^2, a^2)$		
	$a^1 = 1$	$a^1 = 2$		$a^2 = 1$	$a^2 = 2$
$s^1 = 1$	7	4	$s^2 = 3$	4	3
$s^1 = 2$	2	5	$s^2 = 4$	3	5

6. The bounds defining the constraints are $\xi^1 = 5$, $\xi^2 = 3.5$.

We solve the quadratic program [QP3], corresponding to the above data, by using MATLAB and obtain

$$\zeta^* = (1.2941, 0, 0, 1.7059, -0.5882, 0.5882, 0, 0, 0, 1, 0, 0.2941, 0.7059, 0, 0, 0).$$

Note that at ζ^* the objective function value is zero and hence it is the global minimum of [QP3]. We have $x^{1*} = (0, 0, 0, 1)$ and $x^{2*} = (0, 0.2941, 0.7059, 0)$ then from Theorem 3 (b) the Nash equilibrium (f^{1*}, f^{2*}) of constrained stochastic game G_{ea}^c , where

$$f^{1*} = ((\alpha, 1 - \alpha), (0, 1)) \text{ for all } \alpha \in [0, 1]$$

$$f^{2*} = ((0, 1), (1, 0))$$

Note that under f^{1*} player 1 can use any randomized strategy at state 1 which comes from the fact that state 1 is transient under f^{1*} . The costs of both the players at Nash equilibrium (f^{1*}, f^{2*}) are

$$C_{ea}^1(f^{1*}, f^{2*}) = 1.2941$$

$$C_{ea}^2(f^{1*}, f^{2*}) = 1.7059.$$

References

1. E. Altman, S. Sarkar, E. Solan, Constrained Markov games with transition probabilities controlled by a single player, SMCTools 07, Nantes, France, October, 26, 2007.
2. E. Altman, K. Avrachenkov, N. Bonneau, M. Debbah, R. El-Azouzi, D. S. Menasche, Constrained cost-coupled stochastic games with independent state processes, *Operations Research Letters* 36 (2008) 160–164.
3. E. Altman, A. Schwartz, Constrained Markov games: Nash equilibria, *Annals of International Society of Dynamic games* 5 (2000) 303–323.
4. F. Giannessi, E. Tomasin, Nonconvex quadratic programs, linear complementarity problems and integer linear programs, in: R. Conti, A. Ruberti (Eds.), 5th Conference on Optimization Techniques Part I, Vol. 3 of Lecture Notes in Computer Science, Springer Berlin / Heidelberg, 1973, pp. 437–449.
5. J. Hu, J. E. Mitchell, J.-S. Pang, An lpcc approach to nonconvex quadratic programs, *Mathematical Programming* (2010) 1–35.
6. A. Hordijk, L. C. M. Kallenberg, Linear programming and Markov games II, in: O. Moeschin, D. Pallaschke (Eds.), *Game Theory and Mathematical Economics*, North-Holland, 1981, pp. 307–320.
7. E. Altman, K. Avrachenkov, R. Marquez, G. Miller, Zero-sum constrained stochastic games with independent state processes, *Mathematical Methods of Operations Research* 62 (2005) 375–386.
8. O. L. Mangasarian, H. Stone, Two-person nonzero-sum games and quadratic programming, *Journal of Mathematical Analysis and Applications* 9 (1964) 348–355.
9. J. A. Filar, T. A. Schultz, F. Thuijsman, O. J. Vrieze, Nonlinear programming and stationary equilibria in stochastic games, *Mathematical Programming* 50 (1991) 227–237.
10. J. Filar, K. Vrieze, *Competitive Markov Decision Processes*, Springer, New York, 1997.
11. T. E. S. Raghavan, J. A. Filar, Algorithms for stochastic games: A survey, *Mathematical Methods of Operations Research* 35 (6) (1991) 437–472.
12. T. Parthasarathy, T. E. S. Raghavan, An order field property for stochastic games when one player controls transition probabilities, *Journal of Optimization Theory and Applications* 33 (1981) 375–392.
13. O. J. Vrieze, Linear programming and undiscounted stochastic games in which one player controls transitions, *OR Spektrum* 3 (1981) 29–35.
14. J. A. Filar, Quadratic programming and the single controller stochastic game, *Journal of Mathematical Analysis and Applications* 113 (1986) 136–147.
15. L. S. Shapley, Stochastic games, *Proceedings of National Academy of Science* 39 (1953) 1095–1100.
16. A. Neyman, S. Sorin (Eds.), *Stochastic Games and their Applications*, Springer, Berlin-Heidelberg, 2003.
17. J. Alvarez-Mena, O. Hernandez-Lerma, Existence of Nash equilibria for constrained stochastic games, *Mathematical Methods of Operations Research* 63 (2006) 261–285.
18. A. Hordijk, L. C. M. Kallenberg, Linear programming and Markov games I, in: O. Moeschin, D. Pallaschke (Eds.), *Game Theory and Mathematical Economics*, North-Holland, 1981, pp. 291–305.
19. E. Altman, K. Avrachenkov, N. Bonneau, M. Debbah, R. El-Azouzi, D. S. Menasche, Constrained stochastic games in wireless networks, in: *IEEE GLOBECOM*, 2007.
20. E. Altman, *Constrained Markov Decision Processes*, Chapman and Hall/CRC, London, 1999.
21. R. W. Cottle, W. C. Mylander, Ritter’s cutting plane method for non-convex quadratic programming, *Integer and nonlinear programming* (1970) 257–283.
22. C. A. Burdet, General quadratic programming, Tech. Rep. w.p. 41-71-2, Carnegie-Mellon University (November 1971).
23. P. B. Zwart, Nonlinear programming: Counter-examples to global optimization algorithms proposed by Ritter and Tui, Tech. rep., Washington University, Dept. of Applied Mathematics and Computer Sciences School of Engineering and Applied Science. Report No. Co -1493-32- (1972).

24. D. Bertsimas, J. N. Tsitsiklis, Introduction to Linear Optimization, Athena Scientific, Massachusetts, 1997.
25. M. Bazaraa, H. Sherali, C. Shetty, Nonlinear Programming Theory and Algorithms, John Wiley and Sons, Inc., U.S.A, Third ed., 2006.

Appendix A

A single mathematical program for average and discounted cost criteria model

The mathematical programs [MP1] and [MP2] that characterize the stationary Nash equilibria of single controller constrained stochastic games with average and discounted cost criteria respectively can be recovered from one mathematical program [MP4] given below.

$$\begin{aligned}
 \text{[MP4]} \quad & \min_{\eta} \left[\left(f^T C^1 x - \left(\mathbf{1}^T z - (\delta^1)^T \xi^1 \right) \right) + \left(f^T C^2 x - \left(v + (1 - \beta) \gamma^T u - (\delta^2)^T \xi^2 \right) \right) \right] \\
 \text{s.t.} \quad & \\
 (i) \quad & v + u(s) \leq \left[(f(s))^T C^2(s) \right]_{a^2} + \sum_{l=1}^{n_2} \delta_l^2 \left[(f(s))^T D^{2,l}(s) \right]_{a^2} \\
 & \quad + \beta \sum_{s' \in S} p(s'|s, a^2) u(s'), \quad \forall s \in S, a^2 \in A^2(s) \\
 (ii) \quad & z(s) \leq \left[C^1(s) x(s) \right]_{a^1} + \sum_{k=1}^{n_1} \delta_k^1 d_{sub}^{1,k}(s, a^1), \quad \forall s \in S, a^1 \in A^1(s) \\
 (iii) \quad & \sum_{(s, a^2) \in \kappa^2} [\delta(s, s') - \beta p(s'|s, a^2)] x(s, a^2) = (1 - \beta) \gamma(s'), \quad \forall s' \in S \\
 (iv) \quad & \sum_{(s, a^2) \in \kappa^2} x(s, a^2) = 1 \\
 (v) \quad & \sum_{(s, a^1) \in \kappa^1} d_{sub}^{1,k}(s, a^1) f(s, a^1) \leq \xi_k^1, \quad \forall k = 1, 2, \dots, n_1 \\
 (vi) \quad & \sum_{s \in S} (f(s))^T D^{2,l}(s) x(s) \leq \xi_l^2, \quad \forall l = 1, 2, \dots, n_2 \\
 (vii) \quad & \sum_{a^1 \in A^1(s)} f(s, a^1) = 1, \quad \forall s \in S \\
 (viii) \quad & f(s, a^1) \geq 0, \quad \forall s \in S, a^1 \in A^1(s) \\
 (ix) \quad & x(s, a^2) \geq 0, \quad \forall s \in S, a^2 \in A^2(s) \\
 (x) \quad & \delta_k^1 \geq 0, \quad \forall k = 1, 2, \dots, n_1 \\
 (xi) \quad & \delta_l^2 \geq 0, \quad \forall l = 1, 2, \dots, n_2.
 \end{aligned}$$

The mathematical program [MP1] can be obtained by putting $\beta = 1$ in [MP4]. For discount factor $\beta \in [0, 1)$ the constraint (iv) of [MP4] is redundant because it can be obtained by summing (iii) over all $s' \in S$ and hence the variable v is also redundant. So, by removing constraint (iv) and variable v from [MP4] we obtain [MP2].